

1	2364	100.0	2364	6	AX695547	Sequence
2	2364	100.0	2782	6	AR360019	Sequence
3	2364	100.0	2782	6	AX622837	Sequence
4	2364	100.0	2782	6	AX622837	Sequence
5	2364	100.0	2782	6	AX695546	Sequence
6	2364	100.0	2782	6	AX777761	Sequence
7	2364	100.0	2782	9	HSU47686	Human signa
8	2356	99.7	2364	6	CQ815791	Sequence
9	2356	99.7	2716	9	HA048730	Homo sapien
10	2356	99.7	5167	9	BC065227	Homo sapi
11	2096.8	88.7	2414	4	AF135123	Scrof
12	1799.2	88.0	2485	4	BT4237933	Bos tauru
13	1991.2	84.2	2671	10	RNST45BG	R.norvegicu
14	1981.6	83.8	4893	10	BC024319	Mus muscu
15	1980	83.8	2940	10	AY040231	Mus muscu
16	1980	83.8	2940	10	AY044901	Mus muscu
17	1980	83.8	2940	10	AY044902	Mus muscu
18	1978.4	83.7	2940	6	AR359993	Sequence
19	1978.4	83.7	2940	6	AX622811	Sequence

QY 241 TTTTACTGAAGATCAAGCTGGGCACATATGCCACACAGCTCCAGAAACACGTATGACCGC 300
Db 241 TTTTACTGAAGATCAAGCTGGGCACATATGCCACACAGCTCCAGAAACACGTATGACCGC 300
QY 301 TGCCCCATGGAGCTGGTTCGGCTGCATCCGCCCATATATTGTACAAATGAACAGAGGTTGGTC 360
Db 301 TGCCCCATGGAGCTGGTTCGGCTGCATCCGCCCATATATTGTACAAATGAACAGAGGTTGGTC 360
QY 361 CGAAGGCCAAACAATGGTAGTCTCCAGCTGGAAGCCTTGTGTATGCCATATCCAGAAA 420
Db 361 CGAAGGCCAAACAATGGTAGTCTCCAGCTGGAAGCCTTGTGTATGCCATATCCAGAAA 420
QY 421 CACCTCCAGATCAACACAGCTTTGAGAGCTGGACCTTTCATCCAGTACACGAGACACAGAAAT 480
Db 421 CACCTCCAGATCAACACAGCTTTGAGAGCTGGACCTTTCATCCAGTACACGAGACACAGAAAT 480
QY 481 GAGTTAAAAAGCTGCACAGACTCAGGAGTACTTTCATCCAGTACACGAGAGACCTG 540
Db 481 GAGTTAAAAAGCTGCACAGACTCAGGAGTACTTTCATCCAGTACACGAGAGACCTG 540
QY 541 AGGATCCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGAGCGCTCTGAGCCGG 600
Db 541 AGGATCCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGAGCGCTCTGAGCCGG 600
QY 601 GAGAGCGCCCTCCAGCAGAACGNGTGTCTCTGGAGGCTGGTTGCAGCGTGAAGCAG 660
Db 601 GAGAGCGCCCTCCAGCAGAACGNGTGTCTCTGGAGGCTGGTTGCAGCGTGAAGCAG 660
QY 661 ACACCTGACAGTACCGGCTGGAGCTGCCCGAGAGACACAGAGACCCCTGCAGCTGCTG 720
Db 661 ACACCTGACAGTACCGGCTGGAGCTGCCCGAGAGACACAGAGACCCCTGCAGCTGCTG 720
QY 721 CGGAAGCAGCAGACCATCATCTGTGATGACAGCTGATCCAGTGAAGCGCGGAGCAG 780
Db 721 CGGAAGCAGCAGACCATCATCTGTGATGACAGCTGATCCAGTGAAGCGCGGAGCAG 780
QY 781 CTGCGCGGGAACGGCGGCCCCCGAGGGCAGCTTGAAGCGTGTACAGCTCTGGTGTGAG 840
Db 781 CTGCGCGGGAACGGCGGCCCCCGAGGGCAGCTTGAAGCGTGTACAGCTCTGGTGTGAG 840
QY 841 AAGTTGGCGGAGATCATCTGGCAGAAACCGGCAGAGATCCGACGGGCTGAGCACCCTGCG 900
Db 841 AAGTTGGCGGAGATCATCTGGCAGAAACCGGCAGAGATCCGACGGGCTGAGCACCCTGCG 900
QY 901 CAGCAGCTGCCATCCCGGCCAGTGGAGAGATGCTGGCCGAGGTCACCGCACCACATC 960
Db 901 CAGCAGCTGCCATCCCGGCCAGTGGAGAGATGCTGGCCGAGGTCACCGCACCACATC 960
QY 961 ACGGACATATCTCAGCCCTGGTGACACAGCAGCTTCATATGAGAGCAGCCCTCTCAG 1020
Db 961 ACGGACATATCTCAGCCCTGGTGACACAGCAGCTTCATATGAGAGCAGCCCTCTCAG 1020
QY 1021 GTCTTGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCCCTGCTGGTGGCGGGAAGCTG 1080
Db 1021 GTCTTGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCCCTGCTGGTGGCGGGAAGCTG 1080
QY 1081 AACGTGCAATGAACCCCGCCAGGTGAAGGCCACCATCATCAGTGAAGCAGCAGGCCAAG 1140
Db 1081 AACGTGCAATGAACCCCGCCAGGTGAAGGCCACCATCATCAGTGAAGCAGCAGGCCAAG 1140
QY 1141 TCTCTGCTCAAGAACGAGAAACCCCGCAATGATTACAGTGGCGGAGATCTTGAACAACCTGC 1200
Db 1141 TCTCTGCTCAAGAACGAGAAACCCCGCAATGATTACAGTGGCGGAGATCTTGAACAACCTGC 1200
QY 1201 TGCGTCAATGAGTACCAACAAGCCACAGGACCCCTTAGTGCCCACTTCAGGAAATATGTCC 1260
Db 1201 TGCGTCAATGAGTACCAACAAGCCACAGGACCCCTTAGTGCCCACTTCAGGAAATATGTCC 1260
QY 1261 CTGAAAACGAATTAAGAGGTCAGACCGTCTGGGGCAGAGTCGGTGACAGAAAGAAAATTT 1320
Db 1261 CTGAAAACGAATTAAGAGGTCAGACCGTCTGGGGCAGAGTCGGTGACAGAAAGAAAATTT 1320

QY 1321 ACAAATCTGTTTGAATCCCAAGTTTCAGTGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1380
Db 1321 ACAAATCTGTTTGAATCCCAAGTTTCAGTGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1380
QY 1381 ACCGTGTCCTTCGCCAGTGGTGTGATCGTTTCATGGCAGCCAGGACAAACAATGCGACGGCC 1440
Db 1381 ACCGTGTCCTTCGCCAGTGGTGTGATCGTTTCATGGCAGCCAGGACAAACAATGCGACGGCC 1440
QY 1441 ACTGTTCTCTGGGCAAAATGCTTTTGCAGAGCTGGCAGGGTGCATTTTGCCTGCTGAC 1500
Db 1441 ACTGTTCTCTGGGCAAAATGCTTTTGCAGAGCTGGCAGGGTGCATTTTGCCTGCTGAC 1500
QY 1501 AAAGTGTCTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGAGCCGAAAGTGAG 1560
Db 1501 AAAGTGTCTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGAGCCGAAAGTGAG 1560
QY 1561 AGCAACCGGGCCCTGACCAAGGAGAACCTCGTGTTCCTGGCGCAGAAACTGTTCAACAAC 1620
Db 1561 AGCAACCGGGCCCTGACCAAGGAGAACCTCGTGTTCCTGGCGCAGAAACTGTTCAACAAC 1620
QY 1621 AGCAGCAGCCACTGGAGGACTACAGTGGCGCTGTCTGTCTCTGCTGCCAGTCCCAAGG 1680
Db 1621 AGCAGCAGCCACTGGAGGACTACAGTGGCGCTGTCTGTCTCTGCTGCCAGTCCCAAGG 1680
QY 1681 GAGAAATTTACAGGACGGAATTTACACTTCTGGCAATGGTTTGACGGTGTGATGGAAGTG 1740
Db 1681 GAGAAATTTACAGGACGGAATTTACACTTCTGGCAATGGTTTGACGGTGTGATGGAAGTG 1740
QY 1741 TTAAAAAACAATCTCAAGCCCTCATTTGGAATGATGGGGCCATTTTGGGGTGTGTAACAAG 1800
Db 1741 TTAAAAAACAATCTCAAGCCCTCATTTGGAATGATGGGGCCATTTTGGGGTGTGTAACAAG 1800
QY 1801 CAACAGCGCCATGACCTACTGATTAACAAGCAGATGGGACCTTCCTCTCGAGATTCAGT 1860
Db 1801 CAACAGCGCCATGACCTACTGATTAACAAGCAGATGGGACCTTCCTCTCGAGATTCAGT 1860
QY 1861 GACTCAGAAATGGCGGCATCACCATTGCTTGGAAATTTGATTTCTCAGGAAAGAAATGTTT 1920
Db 1861 GACTCAGAAATGGCGGCATCACCATTGCTTGGAAATTTGATTTCTCAGGAAAGAAATGTTT 1920
QY 1921 TGGAAATCTGATGCTTTTACCACAGAGATCTTCCATCAGGTCCCTAGCGCAGCGCTTG 1980
Db 1921 TGGAAATCTGATGCTTTTACCACAGAGATCTTCCATCAGGTCCCTAGCGCAGCGCTTG 1980
QY 1981 GGAGACTTGAATTAACCTTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATCTCC 2040
Db 1981 GGAGACTTGAATTAACCTTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATCTCC 2040
QY 2041 AAATACTACACACAGTTCCCTGCGAGTCTGCTAAAGCTGTAAAGCTGTGATGGATACGTG 2100
Db 2041 AAATACTACACACAGTTCCCTGCGAGTCTGCTAAAGCTGTGATGGATACGTG 2100
QY 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCGCGGGGC 2160
Db 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCGCGGGGC 2160
QY 2161 GGCAGCGCAGTATGAGCAGGCCCCCTCCAGCTGTGTGTCCAGGCTCACTAT 2220
Db 2161 GGCAGCGCAGTATGAGCAGGCCCCCTCCAGCTGTGTGTGTCCAGGCTCACTAT 2220
QY 2221 AACATGTACCCACAGAACCCCTGACTCAGTCTTGACACCGATGGGACTTTCGATCTGGAG 2280
Db 2221 AACATGTACCCACAGAACCCCTGACTCAGTCTTGACACCGATGGGACTTTCGATCTGGAG 2280
QY 2281 GACAAATGGAACGTAGCGCGCGTGTGGAGAGCTCCTGGGCGGGCCAAATGGACAGTCAAG 2340
Db 2281 GACAAATGGAACGTAGCGCGCGTGTGGAGAGCTCCTGGGCGGGCCAAATGGACAGTCAAG 2340
QY 2341 TGGATCCCGCAGCAGCAATCGTGA 2364
Db 2341 TGGATCCCGCAGCAGCAATCGTGA 2364

RESULT 2	AR360019	2782 bp	DNA	linear	PAT 17-AUG-2003
LOCUS	Sequence 158 from patent US 6596036.				
DEFINITION	AR360019				
ACCESSION	AR360019				
VERSION	AR360019.1	GI:33766863			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 2782)				
AUTHORS	Derksen,G.C.H., Van Beek,T.A., De Groot,A. and Capelle,A.				
TITLE	Method for the production of a dye preparation based on madder root				
JOURNAL	Patent: US 6596036-A 158 22-JUL-2003;				
FEATURES	Location/Qualifiers				
source	1..2782				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
	Query Match	100.0%;	Score 2364;	DB 6;	Length 2782;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 2364;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGCTGTGTGGATACAAAGCTCAGCAGCTCCAAGGAGAAGCCCTTCATCAGATCGAAGCG	60		
Db	147	ATGCTGTGTGGATACAAAGCTCAGCAGCTCCAAGGAGAAGCCCTTCATCAGATCGAAGCG	206		
Qy	61	TTATATGCCAGCATTTTCCCATTTGAGTGGCGCATTTATTCATCCAGTGGATTGAAGC	120		
Db	207	TTATATGCCAGCATTTTCCCATTTGAGTGGCGCATTTATTCATCCAGTGGATTGAAGC	266		
Qy	121	CAAGCATGGGACTCAGTAGACTTTGTAATCCACAGGAGAACATTAAGGCCACCCAGCTC	180		
Db	267	CAAGCATGGGACTCAGTAGACTTTGTAATCCACAGGAGAACATTAAGGCCACCCAGCTC	326		
Qy	181	CTGAGAGGCCCTGTGTCAGGAGCTGCAGAGAAGGCAGAGCACAGGTGGGGGAAGATGGG	240		
Db	327	CTGAGAGGCCCTGTGTCAGGAGCTGCAGAGAAGGCAGAGCACAGGTGGGGGAAGATGGG	386		
Qy	241	TTTTTTACTGAGATCAAGCTGGGGCAGCTATGSCCAGACAGCTCCAGAACACAGTATGACCGC	300		
Db	387	TTTTTTACTGAGATCAAGCTGGGGCAGCTATGSCCAGACAGCTCCAGAACACAGTATGACCGC	446		
Qy	301	TGCCCATGGAGCTGGTCCGCTGCATCCGCCATATTTGTACAATGAACAGAGGTTGGTC	360		
Db	447	TGCCCATGGAGCTGGTCCGCTGCATCCGCCATATTTGTACAATGAACAGAGGTTGGTC	506		
Qy	361	CGAAGAGCCCAACATGTTAGCTTCCAGCTGGAAAGCCTTGTGATGCCATGTCTCCAGAAA	420		
Db	507	CGAAGAGCCCAACATGTTAGCTTCCAGCTGGAAAGCCTTGTGATGCCATGTCTCCAGAAA	566		
Qy	421	CACCTCCAGATCAACAGAGCTTTGAGAGCTCGAGCTGGTTCACGACAGGACACAGAGAAAT	480		
Db	567	CACCTCCAGATCAACAGAGCTTTGAGAGCTCGAGCTGGTTCACGACAGGACACAGAGAAAT	626		
Qy	481	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCTCAGTACACAGGAGAGCCTG	540		
Db	627	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCTCAGTACACAGGAGAGCCTG	686		
Qy	541	AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCCGAGCTGAGCCCGCCAGGAGCGTCTGAGCGG	600		
Db	687	AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCCGAGCTGAGCCCGCCAGGAGCGTCTGAGCGG	746		
Qy	601	GAGAGCGCCCTCCAGCAGAACAGGTGTCTCTCGAGGCGCTGGTTGCAGCGTGAGGCACAG	660		
Db	747	GAGAGCGCCCTCCAGCAGAACAGGTGTCTCTCGAGGCGCTGGTTGCAGCGTGAGGCACAG	806		
Qy	661	ACACTGCAGCATGACCGCGTGGAGCTGCCCGAGAGGACACCAAGAGACCCCTCAGCTGCTG	720		
Db	807	ACACTGCAGCATGACCGCGTGGAGCTGCCCGAGAGGACACCAAGAGACCCCTCAGCTGCTG	866		
Qy	721	CGGAAGCAGCAGACCACTCATCTCTGGATGACAGAGCTGATCTCAGTGGAGCGCGCAGCAG	780		

Db 1947 CAACAGGCCCATGACCTACTGATTAAACAGCCAGATGGACCTTCTCTCCTCAGATTTCAGT 2006

Qy 1861 GACTCAGAAATGGCGGATCACCATTGCTGTGGAAGTTTGATTTCTCAGGAAGAATGTTT 1920

Db 2007 GACTCAGAAATGGCGGATCACCATTGCTGTGGAAGTTTGATTTCTCAGGAAGAATGTTT 2066

Qy 1921 TGGATCTGATGCTTTTACACACAGAGACTTCTCCATCAGGTCCCTAGCGCGCTTG 1980

Db 2067 TGGAACTGATGCTTTTACACACAGAGACTTCTCCATCAGGTCCCTAGCGCGCTTG 2126

Qy 1981 GGAGACTTGAATTACCTTATCTAGCTGTTTCTGATCGGCCAAAAGATGAAGTATACTCC 2040

Db 2127 GGAGACTTGAATTACCTTATCTAGCTGTTTCTGATCGGCCAAAAGATGAAGTATACTCC 2186

Qy 2041 AAATACTACACACAGTTCCCTGGAGTCTGCTACTGCTAAAGCTGTTGATGGAATACGTG 2100

Db 2187 AAATACTACACACAGTTCCCTGGAGTCTGCTACTGCTAAAGCTGTTGATGGAATACGTG 2246

Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2160

Db 2247 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2306

Qy 2161 GGCAGCGCATCTACATGACACAGCCCTCCCGAGCTGTGTGCTCCAGGCTCACTAT 2220

Db 2307 GGCAGCGCATCTACATGACACAGCCCTCCCGAGCTGTGTGCTCCAGGCTCACTAT 2366

Qy 2221 AACATGTACCCACAGAACCTTGACTGCTCTTGACACCGATGGGACTTCGATCTGGAG 2280

Db 2367 AACATGTACCCACAGAACCTTGACTGCTCTTGACACCGATGGGACTTCGATCTGGAG 2426

Qy 2281 GACCAATGAGAGTACCGCGGCTGTGAGAGCTCTGGCGCGCCAAATGGACAGTCAG 2340

Db 2427 GACCAATGAGAGTACCGCGGCTGTGAGAGCTCTGGCGCGCCAAATGGACAGTCAG 2486

Qy 2341 TGGATCCCGCACGACCAATCGTGA 2364

Db 2487 TGGATCCCGCACGACCAATCGTGA 2510

RESULT 3

AX622837

LOCUS AX622837 2782 bp DNA linear PAT 20-FEB-2003

DEFINITION Sequence 158 from Patent WO0224867.

ACCESSION AX622837

VERSION AX622837.1 GI:28450816

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Pedersen,F.S., Sorensen,A.B., Hernandez,J.M., Nielsen,A.A. and Moving,H.O.

AUTHORS Novel compositions and methods for lymphoma and leukemia

TITLE Patent: WO 0224867-A,158 28-MAR-2002;

JOURNAL Aarhus University (DK)

FEATURES Location/Qualifiers

source 1..2782

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 2364; DB 6; Length 2782;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTGTGTGATCAAGACTCAGAGCTCCAGAGAGAGCGCTTCATCAGATGCAAGCG 60

Db 147 ATGGCTGTGTGATCAAGACTCAGAGCTCCAGAGAGAGCGCTTCATCAGATGCAAGCG 206

Qy 61 TTATATGGCCACATTTTCCCATTCAGGTGGGCAATTTATTTATCCAGTGGATTCAGAGC 120

Db 207 TTATATGGCCACGATTTTCCCATTTGAGGTGGGCAATTTATTTATCCAGTGGATTCAGAGC 266

Qy 121 CAAGCATGGGACTCAGTAGATCTTTGATATATCCAGAGAGAACATTAAAGGCCACCAGCTC 180

Db 267 CAAGCATGGGACTCAGTAGATCTTTGATATATCCAGAGAGAACATTAAAGGCCACCAGCTC 326

Qy 181 CTGGAGGGCTGTGTGAGGAGCTGCAGAAAGCGACAGCACAGGTTGGGGGAAGATGGG 240

Db 327 CTGGAGGGCTGTGTGAGGAGCTGCAGAAAGCGACAGCACAGGTTGGGGGAAGATGGG 386

Qy 241 TTTTATCTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACAOTATGACCGC 300

Db 387 TTTTATCTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACAOTATGACCGC 446

Qy 301 TGCCCCATGAGCTGTGCTCGCATCCGCCATATATATTGTATCAATGAACAGAGTTGGTC 360

Db 447 TGCCCCATGAGCTGTGCTCGCATCCGCCATATATATTGTATCAATGAACAGAGTTGGTC 506

Qy 361 CGAGAACCCAAATGTTAGCTCTCCAGCTGGAGCCCTTGTGATGCTCATGTCCCGAATA 420

Db 507 CGAGAACCCAAATGTTAGCTCTCCAGCTGGAGCCCTTGTGATGCTCATGTCCCGAATA 566

Qy 421 CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGCGACTGCTGCTACGACAGGACACAGAAAT 480

Db 567 CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGCGACTGCTGCTACGACAGGACACAGAAAT 626

Qy 481 GAGTTAAAAAGCTGACAGAGACTCAGAGAGTACTTTCAATCATCCAGTACAGAGAGCTG 540

Db 627 GAGTTAAAAAGCTGACAGAGACTCAGAGAGTACTTTCAATCATCCAGTACAGAGAGCTG 686

Qy 541 AGGATCCAAAGCTGAGTTTGGCCGCTGCGCCAGCTGAGCCCGCCAGGAGCGTCTGAGCGG 600

Db 687 AGGATCCAAAGCTGAGTTTGGCCGCTGCGCCAGCTGAGCCCGCCAGGAGCGTCTGAGCGG 746

Qy 601 GAGACGGCCCTCCAGCAGAAAGCAGGTGTCTCTGAGGCGCTGTTGACAGCTGAGGACAG 660

Db 747 GAGACGGCCCTCCAGCAGAAAGCAGGTGTCTCTGAGGCGCTGTTGACAGCTGAGGACAG 806

Qy 661 ACCTGACAGAGTACCGGCTGGAGCTGCCCGAGAGACACAGAAAGACCTTGACAGCTGCTG 720

Db 807 ACCTGACAGAGTACCGGCTGGAGCTGCCCGAGAGACACAGAAAGACCTTGACAGCTGCTG 866

Qy 721 CGGAAGCAGAGACCATCATCTCGGATGACAGCTGATCCAGTGAAGCGGCGCAGCAG 780

Db 867 CGGAAGCAGAGACCATCATCTCGGATGACAGCTGATCCAGTGAAGCGGCGCAGCAG 926

Qy 781 CTGGCCGGAAACGGCGGGCCCCCGAGGCGAGCTTGACAGCTGTCTAGTCTCTGTTGTGAG 840

Db 927 CTGGCCGGAAACGGCGGGCCCCCGAGGCGAGCTTGACAGCTGTCTAGTCTCTGTTGTGAG 986

Qy 841 AAGTTGGCGGAGATCATCTGGCAGAACCGGCGAGAGATCCGACGGCTGAGACCTCTGCG 900

Db 987 AAGTTGGCGGAGATCATCTGGCAGAACCGGCGAGAGATCCGACGGCTGAGACCTCTGCG 1046

Qy 901 CAGCAGCTGCCATCCCGCGCCAGTGGAGAGATGCTGGCCGAGGTCAACGCCACCATC 960

Db 1047 CAGCAGCTGCCATCCCGCGCCAGTGGAGAGATGCTGGCCGAGGTCAACGCCACCATC 1106

Qy 961 ACGGACATTTATCTCAGCCCTGGTGACCGACGCTTCAATTTGAGAAAGAGAGCTCTCTCAG 1020

Db 1107 ACGGACATTTATCTCAGCCCTGGTGACCGACGCTTCAATTTGAGAAAGAGAGCTCTCTCAG 1166

Qy 1021 GTCCTGAAGACCCAGACCAAGTTTGGACCACTGTGGCGCTGCTGGTGGGGGGAAGCTG 1080

Db 1167 GTCCTGAAGACCCAGACCAAGTTTGGACCACTGTGGCGCTGCTGGTGGGGGGAAGCTG 1226

Qy 1081 AAGCTGCACATGAACCCCGCCAGGTGAAGGCCACCATCATCTAGTGAAGCAGCAGGCGAAG 1140

Db 1227 AAGCTGCACATGAACCCCGCCAGGTGAAGGCCACCATCATCTAGTGAAGCAGCAGGCGAAG 1286

Qy 1141 TCTCTGCTCAAGAACGAGAACACCCGCAATGATTTACAGTGGCGAGATCTTGAAACAATGCG 1200

Db 1287 TCTCTGCTCAAGAACGAGAACACCCGCAATGATTTACAGTGGCGAGATCTTGAAACAATGCG 1346

Db 687 AGGATCCAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGACGGCTGAGCCGG 746
Qy 601 GAGACGGCCCTCCAGCAGAAGCAGTGTCTCTGGAGGCTGTGTTGACGCTGAGGCACAG 660
Db 747 GAGACGGCCCTCCAGCAGAAGCAGTGTCTCTGGAGGCTGTGTTGACGCTGAGGCACAG 806
Qy 661 AACTGTCAGCAGTACCGGTCGAGTGTCCCGAGAGAGCAACGAGACCCCTGACGTGTGTG 720
Db 807 AACTGTCAGCAGTACCGGTCGAGTGTCCCGAGAGAGCAACGAGACCCCTGACGTGTGTG 866
Qy 721 CGGAGCAGCAGACCATCATCTGTCGACGAGCTGATCCAGTGTGAGAGCGGCGCAGCAG 780
Db 867 CGGAGCAGCAGACCATCATCTGTCGACGAGCTGATCCAGTGTGAGAGCGGCGCAGCAG 926
Qy 781 CTGCGCGGGAACGGCGGCCCCCGAGGCGAGCCCTGGACGTGCTACAGTCTCTGTGTGAG 840
Db 927 CTGCGCGGGAACGGCGGCCCCCGAGGCGAGCCCTGGAGCTGTACAGTCTCTGTGTGAG 986
Qy 841 AAGTTGGCGGAGATCATCTGTCGACGAGCTGATCCAGTGTGAGAGCGGCGCAGCAGCTCTGC 900
Db 987 AAGTTGGCGGAGATCATCTGTCGAGAACCGGCGAGCAGATCCGCGAGGCTGTACAGCTCTCTGC 1046
Qy 901 CAGCAGCTGCCCATCCCCCGGCCAGTGGAGGAGATGCTGGCGGAGTCAACGCCACCATC 960
Db 1047 CAGCAGCTGCCCATCCCCCGGCCAGTGGAGGAGATGCTGGCGGAGTCAACGCCACCATC 1106
Qy 961 ACGGACATTAATCTCAGCCCTGGTGACCAAGCGTTTCATTAATGAGAGCAGCCCTCTCAG 1020
Db 1107 ACGGACATTAATCTCAGCCCTGGTGACCAAGCGTTTCATTAATGAGAGCAGCCCTCTCAG 1166
Qy 1021 GTCTGAAGACCCAGACCAAGTTTGCAGCCTGTGTGCGCTGTGTGGCGGGAAGCTG 1080
Db 1167 GTCTGAAGACCCAGACCAAGTTTGCAGCCTGTGTGCGCTGTGTGGCGGGAAGCTG 1226
Qy 1081 AAGTGCACATGAACCCCCCAGGTGAAGCCACCATCATCAGTGAAGCAGCGCCCAAG 1140
Db 1227 AAGTGCACATGAACCCCCCAGGTGAAGCCACCATCATCAGTGAAGCAGCGCCCAAG 1286
Qy 1141 TCTGTCTCAAGAACCGAGAACACCGCAATGATTACAGTGGCGAGATCTTGAACAACCTGC 1200
Db 1287 TCTGTCTCAAGAACCGAGAACACCGCAATGATTACAGTGGCGAGATCTTGAACAACCTGC 1346
Qy 1201 TGCGTCATGGAGTACCAACAGCCACAGGCAACCTTAGTGCCCATTCAGGAATATGTCC 1260
Db 1347 TGCGTCATGGAGTACCAACAGCCACAGGCAACCTTAGTGCCCATTCAGGAATATGTCC 1406
Qy 1261 CTGAAACGAATTAAGAGGTGACAGCGTGTGGGCGAGTGGTGAAGAGAAAAATTT 1320
Db 1407 CTGAAACGAATTAAGAGGTGACAGCGTGTGGGCGAGTGGTGAAGAGAAAAATTT 1466
Qy 1321 ACAATCTCTGTTGAATCCAGTTCAGTGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1380
Db 1467 ACAATCTCTGTTGAATCCAGTTCAGTGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1526
Qy 1381 ACCCTGCTCCCTGCGAGTGGTGTGATGTTTCATGGCAGCAGGACCAATGCGACGGCC 1440
Db 1527 ACCCTGCTCCCTGCGAGTGGTGTGATGTTTCATGGCAGCAGGACCAATGCGACGGCC 1586
Qy 1441 ACTGTCTCTGGGCAATGCTTTTTCAGAGGCTGCGAGGGTGCCATTTGCGTGCCTTGAC 1500
Db 1587 ACTGTCTCTGGGCAATGCTTTTTCAGAGGCTGCGAGGGTGCCATTTGCGTGCCTTGAC 1646
Qy 1501 AAGTGTCTGGCCACAGCTGTGAGCGCTCAACATGAAATTTCAAGGCGGAAGTGCAG 1560
Db 1647 AAGTGTCTGGCCACAGCTGTGAGCGCTCAACATGAAATTTCAAGGCGGAAGTGCAG 1706
Qy 1561 AGCAACGGGCGCTGACCAAGGAGAACCTCTGTGTTCTGGCGGAGAACTGTTCAACAAC 1620
Db 1707 AGCAACGGGCGCTGACCAAGGAGAACCTCTGTGTTCTGGCGGAGAACTGTTCAACAAC 1766
Qy 1621 AGCAGCAGCCACCTGGAGGACTACAGTGGCCCTGTCTGTCTCTGCTGCCAGTTCACAGG 1680

Db 1767 AGCAGCAGCCACCTGGAGGACTACAGTGGCCCTGTCTGTGTCTCTGTCCTGTCACAGG 1826
Qy 1681 GAGAATTTTACAGGACCGGAATTAACACTTTCTTGGCAATGGTTTGGACGGTGTGATGGAAGTG 1740
Db 1827 GAGAATTTTACAGGACCGGAATTAACACTTTCTTGGCAATGGTTTGGACGGTGTGATGGAAGTG 1886
Qy 1741 TTAATAAACATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTGTGTAACAAG 1800
Db 1887 TTAATAAACATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTGTGTAACAAG 1946
Qy 1801 CAACAGGCCCATGACTACTTAAACAAGCCAGATGGGACCTTCTCTCAGATTTCAAGT 1860
Db 1947 CAACAGGCCCATGACTACTTAAACAAGCCAGATGGGACCTTCTCTCAGATTTCAAGT 2006
Qy 1861 GACTCAGAAATTTGGCGCATCAACATTCCTTTGGAAAGTTTGAATCTCAGGAAAGAAATGTTT 1920
Db 2007 GACTCAGAAATTTGGCGCATCAACATTCCTTTGGAAAGTTTGAATCTCAGGAAAGAAATGTTT 2066
Qy 1921 TGGAACTGTATGCTTTTAAACACGAGACTTCTCCATCAGTCCCTAGCCGACCGCTTG 1980
Db 2067 TGGAACTGTATGCTTTTAAACACGAGACTTCTCCATCAGTCCCTAGCCGACCGCTTG 2126
Qy 1981 GGAGACTTGAATTAACCTTATCTACGTGTTTCTGATCGGCCAAAGATGAAGTATATCTCC 2040
Db 2127 GGAGACTTGAATTAACCTTATCTACGTGTTTCTGATCGGCCAAAGATGAAGTATATCTCC 2186
Qy 2041 AAATATACACACAGTTCCCTGCGAGTCTGCTACTCTAAAGCTGTGTGATGGATACGTG 2100
Db 2187 AAATATACACACAGTTCCCTGCGAGTCTGCTACTCTAAAGCTGTGTGATGGATACGTG 2246
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCTTCTGATCGGCCAAAGATGAAGTATATCTCC 2160
Db 2247 AAGCCACAGATCAAGCAAGTGGTCTTCTGATCGGCCAAAGATGAAGTATATCTCC 2306
Qy 2161 GGACGCCACAGTACATGAGACCGCCCTCCCGAGCTGTGTCTCCAGGCTCACTAT 2220
Db 2307 GGACGCCACAGTACATGAGACCGCCCTCCCGAGCTGTGTCTCCAGGCTCACTAT 2366
Qy 2221 AACATGTACCACAGAACCTGACTCAGTCTTGTGACACCGATGGGAGCTTCGATCTGGAG 2280
Db 2367 AACATGTACCACAGAACCTGACTCAGTCTTGTGACACCGATGGGAGCTTCGATCTGGAG 2426
Qy 2281 GACCAATGAGCTGAGCGCGCTGTGAGGAGCTCTGGCGCGCAATGGACAGTCAAG 2340
Db 2427 GACCAATGAGCTGAGCGCGCTGTGAGGAGCTCTGGCGCGCAATGGACAGTCAAG 2486
Qy 2341 TGGATCCCGCAGCACAATCGTGA 2364
Db 2487 TGGATCCCGCAGCACAATCGTGA 2510

RESULT 5
LOCUS AX777761 2782 bp DNA linear PAT 14-JUL-2003
DEFINITION Sequence 158 from Patent WO0302726.
ACCESSION AX777761
VERSION AX777761.1 GI:32694773
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Pedersen,F.S., Sorensen,A.B., Hernandez,J.M., Nielsen,A.A. and
Moving,H.
TITLE Novel compositions and methods for diagnosis and treatment of
lymphoma and leukemia
JOURNAL Patent: WO 0302726-A 158 03-APR-2003;
UNIVERSITY OF AARHUS (DK)
FEATURES
source Location/Qualifiers
1 .2782
/organism="Homo sapiens"
/mol_type="unassigned DNA"

ORIGIN				/db_xref="taxon:9606"			
Query Match				100.0%; Score 2364; DB 6; Length 2782;			
Best Local Similarity				100.0%; Pred. No. 0;			
Matches 2364; Conservative				0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	ATGGCTGTGTGATCAAGCTCAGCACTCAGAGCTCAGAGGAGAGCCCTTTCATCAGATGCAAGCG	60				
Db	147	ATGGCTGTGTGATCAAGCTCAGCACTCAGAGCTCAGAGGAGAGCCCTTTCATCAGATGCAAGCG	206				
Qy	61	TTATATGGCAGCATTTTCCCATTTGAGGTGGCGCATTTATCCAGTGGATTTGAAGC	120				
Db	207	TTATATGGCAGCATTTTCCCATTTGAGGTGGCGCATTTATTTATCCAGTGGATTTGAAGC	266				
Qy	121	CAAGCATGGGACTCAGTAGACTTTGATATTCACAGGAGAAATTAAGGCCACCCAGCTC	180				
Db	267	CAAGCATGGGACTCAGTAGACTTTGATATTCACAGGAGAAATTAAGGCCACCCAGCTC	326				
Qy	181	CTGGAGGGCTGTGTGAGGAGCTGCAGAAAGGAGCAGCACAGGTGGGGGAAGATGGG	240				
Db	327	CTGGAGGGCTGTGTGAGGAGCTGCAGAAAGGAGCAGCACAGGTGGGGGAAGATGGG	386				
Qy	241	TTTTTACTGAAGATCAAGCTGGGGCACTATGCGACACAGCTCCAGAACACGATGACCGC	300				
Db	387	TTTTTACTGAAGATCAAGCTGGGGCACTATGCGACACAGCTCCAGAACACGATGACCGC	446				
Qy	301	TGCCCCATGGAGTGGTGGCTGGCTCGATCCGCCATATATTGTAACAAGACAGAGTTGGTC	360				
Db	447	TGCCCCATGGAGTGGTGGCTGGCTCGATCCGCCATATATTGTAACAAGACAGAGTTGGTC	506				
Qy	361	CGAAGAGCCAAATGGTAGCTCTCCAGCTGGAAGCTTGTGATGCCATGTCCAGAAA	420				
Db	507	CGAAGAGCCAAATGGTAGCTCTCCAGCTGGAAGCTTGTGATGCCATGTCCAGAAA	566				
Qy	421	CACCTCCAGATCAACACAGCTTTGAGGAGCTGCGACTGGTCAACGAGGACACAGAGAT	480				
Db	567	CACCTCCAGATCAACACAGCTTTGAGGAGCTGCGACTGGTCAACGAGGACACAGAGAT	626				
Qy	481	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTCATCATCCAGTACCAGGAGCCCTG	540				
Db	627	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTCATCATCCAGTACCAGGAGCCCTG	686				
Qy	541	AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCGCAGGAGCGTCTGAGCCGG	600				
Db	687	AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCGCAGGAGCGTCTGAGCCGG	746				
Qy	601	GAGAGCGCCCTCAGCAGAAAGCAGGTGTCTTGGAGGCTGTGTGAGGCTGAGGCGACAG	660				
Db	747	GAGAGCGCCCTCAGCAGAAAGCAGGTGTCTTGGAGGCTGTGTGAGGCTGAGGCGACAG	806				
Qy	661	ACACTGCAGCAGTACCGCGTGGAGTCCCGAGAGACACAGAGACCTTCAGCTGCTG	720				
Db	807	ACACTGCAGCAGTACCGCGTGGAGTCCCGAGAGACACAGAGACCTTCAGCTGCTG	866				
Qy	721	CGGAAGCAGCAGACATCATCTGTGATGACAGCTGATCCAGTGAAGCGCGGAGCAG	780				
Db	867	CGGAAGCAGCAGACATCATCTGTGATGACAGCTGATCCAGTGAAGCGCGGAGCAG	926				
Qy	781	CTGGCGGGAAACCGGGGGCCCCCGAGGAGCCTTCAGCTGCTACAGTCTTGGTGTGAG	840				
Db	927	CTGGCGGGAAACCGGGGGCCCCCGAGGAGCCTTCAGCTGCTACAGTCTTGGTGTGAG	986				
Qy	841	AAATTGGCGGAGATCATCTGGCAGAACCGGCGAGCAGATCCGAGGGCTGAGCACCTCTGC	900				
Db	987	AAATTGGCGGAGATCATCTGGCAGAACCGGCGAGCAGATCCGAGGGCTGAGCACCTCTGC	1046				
Qy	901	CAGCAGCTCCCATCCCGGGCCAGTGGAGGAGATGCTGGCCGAGTCAACGCCACCATC	960				
Db	1047	CAGCAGCTCCCATCCCGGGCCAGTGGAGGAGATGCTGGCCGAGTCAACGCCACCATC	1106				
Qy	961	ACGGACATTTATCTCAGCCCTGGTGCACAGCAGCTTCATCATTTGAGAAGCAGCCTCTCTCAG	1020				
Db							
1107	ACGGACATTTATCTCAGCCCTGGTGACACAGCAGCTTCATCATTTGAGAAGCAGCCTCTCTCAG	1166					
Qy	1021	GTCTCTGAAGACCCAGACCAAGTTTGCAGCCTCTGTGCGCTCTGTGTGGCGGGAAGCTG	1080				
Db	1167	GTCTCTGAAGACCCAGACCAAGTTTGCAGCCTCTGTGCGCTCTGTGTGGCGGGAAGCTG	1226				
Qy	1081	AACTGTGCATGAACCCCGGAGGAGGCAACATCATCATCAGTGAGCAGGAGGCAAG	1140				
Db	1227	AACTGTGCATGAACCCCGGAGGAGGCAACATCATCATCAGTGAGCAGGAGGCAAG	1286				
Qy	1141	TCTCTCTCAGAAACAGAAACCCCGCAATGATTAAGTGGCGAGATCTTGAACACTGC	1200				
Db	1287	TCTCTCTCAGAAACAGAAACCCCGCAATGATTAAGTGGCGAGATCTTGAACACTGC	1346				
Qy	1201	TGCGTCTGAGGATACCAACAGCCACAGCAGCCCTTAGTGCCACTTTCAGGAAATGTCC	1260				
Db	1347	TGCGTCTGAGGATACCAACAGCCACAGCAGCCCTTAGTGCCACTTTCAGGAAATGTCC	1406				
Qy	1261	CTGAAACCGAAATTAAGAGGTCAAGCCGTGTGGGGCAGAGTCGGTCAAGAAAAATTT	1320				
Db	1407	CTGAAACCGAAATTAAGAGGTCAAGCCGTGTGGGGCAGAGTCGGTCAAGAAAAATTT	1466				
Qy	1321	ACAATCTCTGTTGAATCCAGTTCAGTCTTGGTGGAAATGAGCTGGTTTTCAGTCAAG	1380				
Db	1467	ACAATCTCTGTTGAATCCAGTTCAGTCTTGGTGGAAATGAGCTGGTTTTCAGTCAAG	1526				
Qy	1381	ACCTGTCTCCCTGCGAGTGTGGTGTATCGTTATGCGCAGCAGGACAAATCGACGGCC	1440				
Db	1527	ACCTGTCTCCCTGCGAGTGTGGTGTATCGTTATGCGCAGCAGGACAAATCGACGGCC	1586				
Qy	1441	ACTGTTCTCTGGGCAATGCTTTTTCAGAGCCTGCGAGGGTGCCATTTGCCGTGCTGAC	1500				
Db	1587	ACTGTTCTCTGGGCAATGCTTTTTCAGAGCCTGCGAGGGTGCCATTTGCCGTGCTGAC	1646				
Qy	1501	AAAGTCTGTGGCAGCAGCTGTGAGGGGCTCAACATGAATTCAGGCCAGAGTGCAG	1560				
Db	1647	AAAGTCTGTGGCAGCAGCTGTGAGGGGCTCAACATGAATTCAGGCCAGAGTGCAG	1706				
Qy	1561	AGCAACCGGGGCTCAGCAAGGAGAACTCTGTGTTCTGGCGCAGAACTGTTCACAAAC	1620				
Db	1707	AGCAACCGGGGCTCAGCAAGGAGAACTCTGTGTTCTGGCGCAGAACTGTTCACAAAC	1766				
Qy	1621	AGCAGCAGCCACTCGAGGAGCTACAGTGGCCCTGTGTCTGTCTGTGTCAGTTCACAGG	1680				
Db	1767	AGCAGCAGCCACTCGAGGAGCTACAGTGGCCCTGTGTCTGTCTGTGTCAGTTCACAGG	1826				
Qy	1681	GAGAAATTTACAGAGCGGAAATTAACACTTTTCGCGAATGGTTTGAAGTGTGATGAGTG	1740				
Db	1827	GAGAAATTTACAGAGCGGAAATTAACACTTTTCGCGAATGGTTTGAAGTGTGATGAGTG	1886				
Qy	1741	TTAAAAAACAATCTCAAGCCTCATTTGGAATGATGGGCGCATTTTGGGGTTTCTAAACAAG	1800				
Db	1887	TTAAAAAACAATCTCAAGCCTCATTTGGAATGATGGGCGCATTTTGGGGTTTCTAAACAAG	1946				
Qy	1801	CAACAGGCCCATGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTCT	1860				
Db	1947	CAACAGGCCCATGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTCT	2006				
Qy	1861	GACTCAGAAATTTGGCGGAGTACCACTTGTGGTGGAAATTTGATCTCAGGAAAGATGTTT	1920				
Db	2007	GACTCAGAAATTTGGCGGAGTACCACTTGTGGTGGAAATTTGATCTCAGGAAAGATGTTT	2066				
Qy	1921	TGGAATCTGATGCTTTTACCAACAGAGACTTCTCATCAGGTCTCCAGCCAGCCTTG	1980				
Db	2067	TGGAATCTGATGCTTTTACCAACAGAGACTTCTCATCAGGTCTCCAGCCAGCCTTG	2126				
Qy	1981	GGAGACTTGAATTAACCTTATCTACCTGTTTCTGATCGGCCAAAAAGATGAACTACTCC	2040				
Db	2127	GGAGACTTGAATTAACCTTATCTACCTGTTTCTGATCGGCCAAAAAGATGAACTACTCC	2186				
Qy	2041	AAATACTACACCAAGTTCCCTGCGAGTCTGCTACTGTCTAAAGCTGTTGATGATGATG	2100				
Db	2187	AAATACTACACCAAGTTCCCTGCGAGTCTGCTACTGTCTAAAGCTGTTGATGATGATG	2246				

QY 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTGTGAAGCATCTGCAGATGCCGGGGC 2160
Db 2247 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTGTGAAGCATCTGCAGATGCCGGGGC 2306
QY 2161 GGCAGCGCCAGTACATGAGCAGCGCCCTCCCGAGCTGTGTGTCCCGAGGCTCACTAT 2220
Db 2307 GGCAGCGCCAGTACATGAGCAGCGCCCTCCCGAGCTGTGTGTCCCGAGGCTCACTAT 2366
QY 2221 AACATGTACCCACAGAACCTTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Db 2367 AACATGTACCCACAGAACCTTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAG 2426
QY 2281 GACACAATGAGTAGCGCGGGGTGTGAGGAGCTCTGGCGCGCCCAATGGACAGTCAG 2340
Db 2427 GACACAATGAGTAGCGCGGGGTGTGAGGAGCTCTGGCGCGCCCAATGGACAGTCAG 2486
QY 2341 TGGATCCCGCAGCACAATCGTGA 2364
Db 2487 TGGATCCCGCAGCACAATCGTGA 2510
RESULT 6
LOCUS HSU47686 2782 bp mRNA linear PRI 23-MAY-1996
DEFINITION Human signal transducer and activator of transcription Stat5B mRNA, complete cds.
ACCESSION U47686
VERSION U47686.1 GI:1330323
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2782)
AUTHORS Lin, J.-X., Mietz, J., Modi, W.S., John, S. and Leonard, W.J.
TITLE Cloning of human Stat5B. Reconstitution of interleukin-2-induced Stat5A and Stat5B DNA binding activity in COS-7 cells
J. Biol. Chem. 271 (18), 10738-10744 (1996)
JOURNAL 96210005
MEDLINE 8631883
PUBMED
REFERENCE 2 (bases 1 to 2782)
AUTHORS Lin, J.-X., Mietz, J., Modi, W.S., John, S. and Leonard, W.J.
TITLE Direct Submission
Submitted (31-JAN-1996) Jian-Xin Lin, Lab of Molecular Immunology, NHLBI, NIH, 9000 Rockville Pike, Bldg. 10, Rm. 7N244, Bethesda, MD 20892-1674, USA
JOURNAL
FEATURES
source
1. .2782
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q11.2"
147..2510
CDS
/note="STAT protein; is activated by IL-2, IL-7, IL-15, growth hormone, IL-3, GM-CSF, thrombopoietin, prolactin, and erythropoietin; tyrosine 699 phosphorylation is required for activation and dimerization of Stat5B"
/product="signal transducer and activator of transcription Stat5B"
/protein_id="AAC50491.1"
/db_xref="GI:1330324"
/translation="MAVWIAQQLQGEALHQMALYGOHFFIEVRHYLSQWIESQAWDSVLDNPENIKATOLLGELVOELOKAEHVGEDGFLKIKLGHYATQLQNTYDRCP MELVRCIRHLYNEORLVREANGSSPAGSLADAMSOKHLOINOTFEELRVTDOTEN ELKLLQQTVEFLIQYSLRIQAQPGPLAQLSPQRLSRALQKQVSLQWLEMLQRE AQLQOYRVEPERGKTLQLLRKQTLILDELQWRRQOLAGNGPFPESLDVLQ SWCKELAEIWNQRIQRIAEHLCOOLPIGPVPEMLAEVNATIDISALVSTFII EKQPPVLTQNTKQTPAATVRLVGGKLVHNNPPOVKATIISEQAKSLKRNENRNDY SGEILNNCCVMYHQATGTLTAHFRNMSLKRIRSDRRGASVTEEFKTLIFESQFSV GGNELVFOVKTLSLPVPVIVHGSQDNNTATVLMDNAFAEPGRVFPFAVPDKVLWPLC

ORIGIN

Query Match 100.0%; Score 2364; DB 9; Length 2782;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTGTGTGGATACAAAGCTCAGCAGCTCAAGGAGAAGCCCTTCATCAGATCAAGCG 60
Db 147 ATGGCTGTGTGGATACAAAGCTCAGCAGCTCAAGGAGAAGCCCTTCATCAGATCAAGCG 206
QY 61 TTATATGCCAGCAGCATTTTCCATTGAGGTGGGCAATTATTATCCCAAGTGAATGAAGC 120
Db 207 TTATATGCCAGCAGCATTTTCCATTGAGGTGGGCAATTATTATCCCAAGTGAATGAAGC 266
QY 121 CAGCATGGGACTCAGTAGATCTTGATAATCCACAGGAGAACATTAAAGGCCACCCAGCTC 180
Db 267 CAGCATGGGACTCAGTAGATCTTGATAATCCACAGGAGAACATTAAAGGCCACCCAGCTC 326
QY 181 CTGAGGGCCCTGGTGCAGGAGCTGCAGAAAGGACAGACACCAAGTGGGGAAAGATGGG 240
Db 327 CTGAGGGCCCTGGTGCAGGAGCTGCAGAAAGGACAGACACCAAGTGGGGAAAGATGGG 386
QY 241 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAACACATATGACCGC 300
Db 387 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAACACATATGACCGC 446
QY 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCATATATTGTACAAATGAACAGAGTTCGTC 360
Db 447 TGCCCCATGGAGCTGGTCCGCTGCATCCGCATATATTGTACAAATGAACAGAGTTCGTC 506
QY 361 CGAGAAGCCAAATAGTGTCTCCAGCTGGAAAGCCTTGCTGATGCCATGCTCCAGAAA 420
Db 507 CGAGAAGCCAAATAGTGTCTCCAGCTGGAAAGCCTTGCTGATGCCATGCTCCAGAAA 566
QY 421 CACTCCAGATCAACACAGACGTTTTGAGAGCTGGACTGTGTACGACGAGGACACAGAGAT 480
Db 567 CACTCCAGATCAACACAGACGTTTTGAGAGCTGGACTGTGTACGACGAGGACACAGAGAT 626
QY 481 GAGTTAAAAGCTGCAGCAGACTCAGAGTACTTTCATCTCAGTACCAAGTACCAAGAGCCTG 540
Db 627 GAGTTAAAAGCTGCAGCAGACTCAGAGTACTTTCATCTCAGTACCAAGTACCAAGAGCCTG 686
QY 541 AGGATCCAAGCTCAGTTTGGCCCGCTGGCCCGCTGAGCCCCCAGAGCGCTCTGAGCCGG 600
Db 687 AGGATCCAAGCTCAGTTTGGCCCGCTGGCCCGCTGAGCCCCCAGAGCGCTCTGAGCCGG 746
QY 601 GAGACGGCCCTCCAGCAGAAAGCAGGTGTCTTGAGAGCGCTGGTTCGACGCTGAGGACAG 660
Db 747 GAGACGGCCCTCCAGCAGAAAGCAGGTGTCTCTGAGAGCGCTGGTTCGACGCTGAGGACAG 806
QY 661 ACATGTCAGCAGTACCCGCTGGAGCTCCCGAGAGCAGACAGAGACCTTCAGCTGCTG 720
Db 807 ACATGTCAGCAGTACCCGCTGGAGCTCCCGAGAGCAGACAGAGACCTTCAGCTGCTG 866
QY 721 CGGAAGCAGCAGACCATCATCTCTGGATCAGCAGCTGATCCAGTGGAAAGCGCGCAGCAG 780
Db 867 CGGAAGCAGCAGACCATCATCTCTGGATCAGCAGCTGATCCAGTGGAAAGCGCGCAGCAG 926
QY 781 CTGCGCGGGAACCGCGGGCCCCCGAGGCGAGCTTCGACGCTGCTGCTGCTGCTGAG 840
Db 927 CTGCGCGGGAACCGCGGGCCCCCGAGGCGAGCTTCGACGCTGCTGCTGCTGCTGAG 986
QY 841 AAGTTGGCGGAGATCATCTGTCAGAACCGGACAGAGATCCGACGGCTGAGACACTCTGC 900
Db 987 AAGTTGGCGGAGATCATCTGTCAGAACCGGACAGAGATCCGACGGCTGAGACACTCTGC 1046
QY 901 CAGCAGCTGCCCTCCCGCGGCCAGTGAGAGATGCTGGCGAGGTCAACGCCCAACATC 960

EALNFKFAEVSQNRGLTKENLVFLAQKLFNNSSHLEDYSGLSVSWSQFNRENLPGR
NYTFWOFDGVMEVLKXHLAPHWDDGAILGFVNKQQAHDILLINKPDTGELLRRPSDSEI
GTTIAKAFDSQERMFNLPFTTRDFSRISLADRLGDLNLIYVFPDRPKDQVYSKY
YTFVPCSAATAKAVDGYVKQIKQVFEFVNASADAGGSATINDQAPSFAVCFQAHY
NMTYPQNPDSVLDTDDGDFLEDTMDVARRVELLGRPMDSQWIPHAQS"

1047 CAGCAGCTCCCATCCCGCCAGTGGAGGAGATGCTGGCCGAGGTCAACGCCACCATC 1106
QY
961 ACGGACATTTATCTCAGCCCTGGTGACGACAGCTTTCATCATTTGAGAGCAGCTCCTCAG 1020
DB
1107 ACGGACATTTATCTCAGCCCTGGTGACGACAGCTTTCATCATTTGAGAGCAGCTCCTCAG 1166
QY
1021 GTCTGAAGACCCAGACCAAGTTTGCAGCCATGTGCGCTCTGCTGGTGGCGGGAAGCTG 1080
DB
1167 GTCTGAAGACCCAGACCAAGTTTGCAGCCATGTGCGCTCTGCTGGTGGCGGGAAGCTG 1226
QY
1081 AAGTGCATGATGACCCCGCCAGTGAAGGCCACCATCATCAGTGAGCAGCAGGCCAG 1140
DB
1227 AAGTGCATGATGACCCCGCCAGTGAAGGCCACCATCATCAGTGAGCAGCAGGCCAG 1286
QY
1141 TCTCTGCTCAAGAAACAGAAACCCCGCAATGATTACAGTGGCGAGATCTTTGAACAACTGC 1200
DB
1287 TCTCTGCTCAAGAAACAGAAACCCCGCAATGATTACAGTGGCGAGATCTTTGAACAACTGC 1346
QY
1201 TCGGTCATGAGTACCAACAGCCACAGGACCCCTTGTAGTGGCCACTTTCAAGGAATATGTC 1260
DB
1347 TCGGTCATGAGTACCAACAGCCACAGGACCCCTTGTAGTGGCCACTTTCAAGGAATATGTC 1406
QY
1261 CTGAACAGAAATTAAGAGGTGACAGCGTCTGGGGCAGAGTGGTGACAGAAAGAAATTT 1320
DB
1407 CTGAACAGAAATTAAGAGGTGACAGCGTCTGGGGCAGAGTGGTGACAGAAAGAAATTT 1466
QY
1321 ACAATCTCTGTTGAATCCAGTTTCAGTGTGGTGAATGAGTGGTGTTCAGGTCAAG 1380
DB
1467 ACATCTCTGTTGAATCCAGTTTCAGTGTGGTGAATGAGTGGTGTTCAGGTCAAG 1526
QY
1381 ACCCTGTCTCCCTGCAAGTGGTGTGATCGTTTCATGCGAGCCAGCAACAATCGACGGCC 1440
DB
1527 ACCCTGTCTCCCTGCAAGTGGTGTGATCGTTTCATGCGAGCCAGCAACAATCGACGGCC 1586
QY
1441 ACTGTTCTCTGGCAATGCTTTTGAGAGCTCTGGAGCCCTGGAGGGTGGCATTTCCGCTGCTGAC 1500
DB
1587 ACTGTTCTCTGGCAATGCTTTTGAGAGCTCTGGAGCCCTGGAGGGTGGCATTTCCGCTGCTGAC 1646
QY
1501 AAGTGTCTGTGGCCAGCTGTGTGAGGCGCTCAACATGAAATTAAGGCCAGCAAGTGCAG 1560
DB
1647 AAGTGTCTGTGGCCAGCTGTGTGAGGCGCTCAACATGAAATTAAGGCCAGCAAGTGCAG 1706
QY
1561 AGCAACCGGGGCTGACCAAGGAGAACCTCGTGTCTCTGGCGCAGAACTGTTCACACAC 1620
DB
1707 AGCAACCGGGGCTGACCAAGGAGAACCTCGTGTCTCTGGCGCAGAACTGTTCACACAC 1766
QY
1621 AGCAGCAGCCACTGAGAGACTACAGTGGCCTGTCTGTCTCTGGTCCAGTTCACACAG 1680
DB
1767 AGCAGCAGCCACTGAGAGACTACAGTGGCCTGTCTGTCTCTGGTCCAGTTCACACAG 1826
QY
1681 GAGAAATTTACAGGACGGAATTAACATTTCTGCAATGTTTTCAGCGTGTGATGGAAGTG 1740
DB
1827 GAGAAATTTACAGGACGGAATTAACATTTCTGCAATGTTTTCAGCGTGTGATGGAAGTG 1886
QY
1741 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTGTCTTAAACAG 1800
DB
1887 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTGTCTTAAACAG 1946
QY
1801 CAACAGGCCCATGACCTACTGATTAACAAGCCAGATGGGACCTTCTCTCTGAGATTCAGT 1860
DB
1947 CAACAGGCCCATGACCTACTGATTAACAAGCCAGATGGGACCTTCTCTCTGAGATTCAGT 2006
QY
1861 GACTCAGAAATTTGGGGCATCACCATTGCTGGAGTTTGAATCTCAGGAAAGATGTTT 1920
DB
2007 GACTCAGAAATTTGGGGCATCACCATTGCTGGAGTTTGAATCTCAGGAAAGATGTTT 2066
QY
1921 TGAATCTGATGCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCCGACCGCTTG 1980
DB
2067 TGAATCTGATGCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCCGACCGCTTG 2126
QY
1981 GGAGACTTGAATTAACCTTATCTACGTTTTCCTGATCGGCCAAAAGATGATATCTCC 2040
DB
2127 GGAGACTTGAATTAACCTTATCTACGTTTTCCTGATCGGCCAAAAGATGATGATATCTCC 2186

QY 2041 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTTTATGATGATACGCTG 2100
DB 2187 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTTTATGATGATACGCTG 2246
QY 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCACTTGCAGATGCGGGGCG 2160
DB 2247 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCACTTGCAGATGCGGGGCG 2306
QY 2161 GGCAGCGCCAGTACATGACACAGCCCTCCCTCCAGCTGTGTCTCCAGGCTCACTAT 2220
DB 2307 GGCAGCGCCAGTACATGACACAGCCCTCCCTCCAGCTGTGTCTCCAGGCTCACTAT 2366
QY 2221 AACATGTATCCCAACAGAACCTGACTCAGTCTTGTGACCCGATGGGACTTCGATCTGGAG 2280
DB 2367 AACATGTATCCCAACAGAACCTGACTCAGTCTTGTGACCCGATGGGACTTCGATCTGGAG 2426
QY 2281 GACAAATGAGAGCTAGCGGCGTGTGAGAGAGCTCTGGGCGGCGCAATGACAGTCAAG 2340
DB 2427 GACAAATGAGAGCTAGCGGCGTGTGAGAGAGCTCTGGGCGGCGCAATGACAGTCAAG 2486
QY 2341 TGGATCCCGCAGCACAATCGTGA 2364
DB 2487 TGGATCCCGCAGCACAATCGTGA 2510

RESULT 7

Q0815791
LOCUS Q0815791 2364 bp DNA linear PAT 03-JUN-2004
DEFINITION Sequence 3 from Patent WO2004042040.
ACCESSION Q0815791
VERSION Q0815791.1 GI:48144323
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Nakachi, H. and Iwama, A.
TITLE Expansion agents for stem cells
JOURNAL Patent: WO 2004042040-A 3 21-MAY-2004;
ReproCELL Inc. (JP)
FEATURES
source
1. .2364
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
1. .2364
/note="unnamed protein product"
/codon_start=1
/protein_id="CAG32954.1"
/db_xref="GI:48144324"

CDS

1
SVDLDNPENIKATQLLEGLVOELQKAEHVQEDGFLKKIKLGHYATQLQNTYDRCP
MELVRCIRHILYNEQRLVREANGSSPAGSLADAMSQKHLQINOTFEELRLVTQDEN
ELKLCQTOEYFIIIOQESLRIQAOFLPGLAOLSPOLRSLETALQOKVSLAWLORE
ACTLOOYVELAEKHOKTOLLRKQOTIILDELQWKERQOLAGNGPPESGLDLVQ
SNCEKLAIHWQRQOIRRAEHLCOOLPIPPVEMLAEVNAITDIISALVTSTFII
EKQPPQVLTQTKFAATVRLVLVGGKLVHMPPOVKATIIISQAKSLLENTRNDY
SGEILNNCCVMYEQATGTLTAHFRNMSLKRIKRRDRRAESVTEKEFTILPESQFSV
GNEILVFOVTKLSLPVVIIVHGSQDNNAATATVLDNFAEPGRVFPVADKLVPLQCL
BALNNKPAEYQSVNRGLTKENLVFLAOKLFNNSSHLEDYSGLSVWSQFNRNLPGR
NYTFQWFDGVMELKHLKPHMNDGAILGFVNKQQAHDLLINKPDGTFLLRFSDEI
GGIITAWKPDQERFMNMPFTTRDFSIIRSLAURLGLNLVLYVFPDRPKDEVSKY
YTPVCEATAKADVGVKPKQKVVPFVNASADAGGSGATIMDOAFSPVCPQAHY
NNYPQNPDSVLDITDGFLEDITMDVARRVEBELLGRPMDSQWIPHAQS"

ORIGIN

Query Match 99.7%; Score 2356; DB 6; Length 2364;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2359; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGGCTGTGTGGATACAAAGCTCAGCAGCTCCAAGGAGAGGCCCTTTCATCAGATGCAAGCG 60

|||||
1 ATGGCTGTGGATACAGCTCAGCAGCTCCAGGAGAGCCCTTCATCAGATGCAGCG 60
QY
61 TTATATGGCAGCATTTTCCCATTTAGGTGGCGCATTTATTTATCCAGTGGATGAAAGC 120
Db
61 TTATATGGCAGCATTTTCCCATTTAGGTGGCGCATTTATTTATCCAGTGGATGAAAGC 120
QY
121 CAAGCATGGGACTCAGTATGATCTTATATCCACAGGAGACATTTAAAGCCACCCAGCTC 180
Db
121 CAAGCATGGGACTCAGTATGATCTTATATCCACAGGAGACATTTAAAGCCACCCAGCTC 180
QY
181 CTGAGGCGCTGGTGCAGGAGCTGCAGAAAGAGGACAGACCCAGTGGCGGAAGATGGG 240
Db
181 CTGAGGCGCTGGTGCAGGAGCTGCAGAAAGAGGACAGACCCAGTGGCGGAAGATGGG 240
QY
241 TTTTCTGAAAGATCAAGCTGGGGCACTATGCCCCACACAGCTCCAGAAACACGTATGACCGC 300
Db
241 TTTTCTGAAAGATCAAGCTGGGGCACTATGCCCCACACAGCTCCAGAAACACGTATGACCGC 300
QY
301 TGCCCCATGGAGCTGGTCCGTGTCATCCGCCATATATTTGTACNATGAACAGAGTTGGTC 360
Db
301 TGCCCCATGGAGCTGGTCCGTGTCATCCGCCATATATTTGTACNATGAACAGAGTTGGTC 360
QY
361 CGAGAAGCCAAACATGTGTAGCTCTCCAGCTGGAAGCCTTGTGTGATGCCATGTCCAGAAA 420
Db
361 CGAGAAGCCAAACATGTGTAGCTCTCCAGCTGGAAGCCTTGTGTGATGCCATGTCCAGAAA 420
QY
421 CACCTCCAGATCAACACAGACGTTTGAGAGCTGCGACTGCGTCA CGCAGGAACACAGAGAA 480
Db
421 CACCTCCAGATCAACACAGACGTTTGAGAGCTGCGACTGCGTCA CGCAGGAACACAGAGAA 480
QY
481 GAGTTTAAAAAGCTGCACAGACTCAGAGTACTTTTCATCATCCAGTACCGAGAGCCCTG 540
Db
481 GAGTTTAAAAAGCTGCACAGACTCAGAGTACTTTTCATCATCCAGTACCGAGAGCCCTG 540
QY
541 AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGCGCTCAGGCCG 600
Db
541 AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGCGCTCAGGCCG 600
QY
601 GAGCGGCCCTCCAGCAGAACAGTGTCTCTGGAGGCTGGTTCGACGCTGAGGACAG 660
Db
601 GAGCGGCCCTCCAGCAGAACAGTGTCTCTGGAGGCTGGTTCGACGCTGAGGACAG 660
QY
661 ACATGCGACGATACCGCTGGAGCTGCCGAGAGCCAGAGACCCCTGCAGCTGCTG 720
Db
661 ACATGCGACGATACCGCTGGAGCTGCCGAGAGCCAGAGACCCCTGCAGCTGCTG 720
QY
721 CGGAAGCAGCAGACCATCATCTGGATGACAGCTGATCCAGTGGAAAGCGCGCAGCAG 780
Db
721 CGGAAGCAGCAGACCATCATCTGGATGACAGCTGATCCAGTGGAAAGCGCGCAGCAG 780
QY
781 CTGGCCGGGAACGGCGGCCCCCGAGGCGAGCTTGAGCGTGCTACATCTGTGTGAG 840
Db
781 CTGGCCGGGAACGGCGGCCCCCGAGGCGAGCTTGAGCGTGCTACATCTGTGTGAG 840
QY
841 AAGTTGGCGGAGATCATCTGGCAGAACCGCAGCAGATCCCGCGGCTGAGCACTCTGC 900
Db
841 AAGTTGGCGGAGATCATCTGGCAGAACCGCAGCAGATCCCGCGGCTGAGCACTCTGC 900
QY
901 CAGCAGCTGCCCATCCCCGGCCAGTGGAGGAGATGCTGGCGGAGTCAAGCGCCACCATC 960
Db
901 CAGCAGCTGCCCATCCCCGGCCAGTGGAGGAGATGCTGGCGGAGTCAAGCGCCACCATC 960
QY
961 ACGGACATTTATCTAGCCCTGGTACACAGCAGTTTCATCATTTAGAGACGACCTCTCAG 1020
Db
961 ACGGACATTTATCTAGCCCTGGTACACAGCAGTTTCATCATTTAGAGACGACCTCTCAG 1020
QY
1021 GTCTGGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCCCTGCTGGTGGCGGGAAGCTG 1080
Db
1021 GTCTGGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCCCTGCTGGTGGCGGGAAGCTG 1080
QY
1081 AACGTGCACATGAACCCCCCGAGGTGAAGGCCACCATCATCAGTGAAGCAGCAGGCCAAG 1140
|||||

Db
1081 AACGTGCACATGAACCCCCCGAGGTGAAGGCCACCATCATCAGTGAAGCAGCAGGCCAAG 1140
QY
1141 TCTCTGCTCAAGAAAGAGAACACCCGCAATGATTTACGTGGCGAGATCTTGAAACAACTGC 1200
Db
1141 TCTCTGCTCAAGAAAGAGAACACCCGCAATGATTTACGTGGCGAGATCTTGAAACAACTGC 1200
QY
1201 TGCGTCACTGGAGTACCAACCAAGCCACAGGACCCCTTAGTGCCCACTTCAGGAATATGTCC 1260
Db
1201 TGCGTCACTGGAGTACCAACCAAGCCACAGGACCCCTTAGTGCCCACTTCAGGAATATGTCC 1260
QY
1261 CTGAAAACGAATTAAGAGGTCAGACCGTGGGGCAGAGTCGGTGACAGAAAGAAAAATTT 1320
Db
1261 CTGAAAACGAATTAAGAGGTCAGACCGTGGGGCAGAGTCGGTGACAGAAAGAAAAATTT 1320
QY
1321 ACAATCTCTGTTTGAATCCAGTTCAGTGTTGGTGGAATAGAGCTGTTTTTCAAGTCAAG 1380
Db
1321 ACAATCTCTGTTTGAATCCAGTTCAGTGTTGGTGGAATAGAGCTGTTTTTCAAGTCAAG 1380
QY
1381 ACCCTGTCCCTGCCAGTGGTGATCGTTTCATGGCAGCCAGGACAAACATGCGACGGCC 1440
Db
1381 ACCCTGTCCCTGCCAGTGGTGATCGTTTCATGGCAGCCAGGACAAACATGCGACGGCC 1440
QY
1441 ACTGTCTCTGGGACAAATGCTTTTGCAGAGCCTTGGCAGGGTGCCATTTGCCGTGCTGAC 1500
Db
1441 ACTGTCTCTGGGACAAATGCTTTTGCAGAGCCTTGGCAGGGTGCCATTTGCCGTGCTGAC 1500
QY
1501 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGAGCCGAAAGTGAG 1560
Db
1501 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGAGCCGAAAGTGAG 1560
QY
1561 AGCAAAACCGGGCCCTGACCAAGGAGAACTCTGTGTTCTGTGGCGCAGAACTGTCAACAAC 1620
Db
1561 AGCAAAACCGGGCCCTGACCAAGGAGAACTCTGTGTTCTGTGGCGCAGAACTGTCAACAAC 1620
QY
1621 AGCAGCAGCCACTGGAGGACTACAGTGGCCCTGTGTGTCTCTGGTCCAGTTCAACAGG 1680
Db
1621 AGCAGCAGCCACTGGAGGACTACAGTGGCCCTGTGTGTCTCTGGTCCAGTTCAACAGG 1680
QY
1681 GAGAAATTTACAGGACCGGAATTTACACTTCTGTGGCAATGGTTTGACGGTGTATGGAAGTG 1740
Db
1681 GAGAAATTTACAGGACCGGAATTTACACTTCTGTGGCAATGGTTTGACGGTGTATGGAAGTG 1740
QY
1741 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGCCATTTTGGGTTTGTAAACAAG 1800
Db
1741 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGCCATTTTGGGTTTGTAAACAAG 1800
QY
1801 CAACAGGCCCATGACCTACTGATTTAAACAGCCAGATGGGACCTTCTCTCAGATTCAGT 1860
Db
1801 CAACAGGCCCATGACCTACTGATTTAAACAGCCAGATGGGACCTTCTCTCAGATTCAGT 1860
QY
1861 GACTCAGAAATGGCGGCATCACCATTGCTTGGAAAGTTTGATTTCTCAGGAAGAAATGTTT 1920
Db
1861 GACTCAGAAATGGCGGCATCACCATTGCTTGGAAAGTTTGATTTCTCAGGAAGAAATGTTT 1920
QY
1921 TGGAAATCTGATGCTTTTACCAACAGAGACTTCTCCATCAGTCCCTAGCCGACCGCTTG 1980
Db
1921 TGGAAATCTGATGCTTTTACCAACAGAGACTTCTCCATCAGTCCCTAGCCGACCGCTTG 1980
QY
1981 GGAGACTTGAATTAACCTTATCTACGTGTTTCTCTGATCGGCCAAAGATGAAGTATCTCC 2040
Db
1981 GGAGACTTGAATTAACCTTATCTACGTGTTTCTCTGATCGGCCAAAGATGAAGTATCTCC 2040
QY
2041 AAATCTACACACAGTTCCCTGGAGTCTGCTACTGCTAAAGCTGTGTGATGGATACGTG 2100
Db
2041 AAATCTACACACAGTTCCCTGGAGTCTGCTACTGCTAAAGCTGTGTGATGGATACGTG 2100
QY
2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCGCGGGGC 2160
Db
2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCGCGGGGC 2160
QY
2161 GGCAGCGCCAGTACATGGAACAGGCCCCCTCCAGCTGTGTGTCCCGCAGGCTCACTAT 2220
Db
2161 GGCAGCGCCAGTACATGGAACAGGCCCCCTCCAGCTGTGTGTGTCCCGCAGGCTCACTAT 2220

Qy 2221 AACATGTACCCACAGAACCTTGACTCAGTCTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Db 2221 AACATGTACCCACAGAACCTTGACTCAGTCTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Qy 2281 GACCAATGGACGTAGCGCGCGTGTGGAGAGCTCTTGGCCCGCCCAATGGACAGTCAG 2340
Db 2281 GACCAATGGACGTAGCGCGCGTGTGGAGAGCTCTTGGCCCGCCCAATGGACAGTCAG 2340
Qy 2341 TGGATCCCGCAGCAGACATCTGA 2364
Db 2341 TGGATCCCGCAGCAGACATCTGA 2364

RESULT 8
HSU48730
LOCUS HSU48730 2716 bp mRNA linear PRI 16-FEB-2000
DEFINITION Homo sapiens transcription factor Stat5b (stat5b) mRNA, complete cds.
ACCESSION U48730
VERSION U48730.2 GI:6981725
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2716)
AUTHORS Silva,C.M., Lu,H. and Day,R.N.
TITLE Characterization and cloning of STAT5 from IM-9 cells and its activation by growth hormone
JOURNAL Mol. Endocrinol. 10 (5), 508-518 (1996)
MEDLINE 96311205
PUBMED 8732682

REFERENCE 2 (bases 1 to 2716)
AUTHORS Silva,C.M. and Lu,H.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1996) Internal Medicine, University of Virginia, Box 511, Charlottesville, VA 22908, USA

REFERENCE 3 (bases 1 to 2716)
AUTHORS Silva,C.M. and Lu,H.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Internal Medicine, University of Virginia, Box 511, Charlottesville, VA 22908, USA
REMARK Sequence update by submitter
COMMENT On Feb 16, 2000 this sequence version replaced gi:1216480.

FEATURES
source

1..2716
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="IM-9"
/cell_type="lymphocyte"
1..2716
/gene="stat5b"
12..2375
/gene="stat5b"
/function="signal transducer and activator of transcription"
/codon_start=1
/product="transcription factor Stat5b"
/protein_id="AAC50485.2"
/db_xref="GI:6981726"

/translation="MAVVIQAQQLQGEALHQMQLYQGFPIEVHRYLSQWIESQAND
SVLDLPENIKATQLLEGLVQELQKAEHQVGEDFLKIKLGHYATQLQNTYDRCP
MELVRCIRHLYNQRLVREANNSSPAGSLADAMSQKHLQINQTFEELRLVTDTEN
ELKKLQOQYEFPIIQYQESLRITQAGPLAQLSPQERLSRETALQQKQVSLQVLEAWLRE
AQTLCQYVELAEKHOKTLLRLKQOTIILDDLIOWKRRQOLAGNGPPGSLDLVQ
SKCEKLAEIWNQEQIPRAHLQCOLPIPGVEMLAEVNAITDIIISALVSTFII
EKQPPVLKQTKFATVRLVLLVGLKLVNHPMPQKATIISEQAKSLIKNENTNDY
SGEILNCCVMEYHQATVTLGAHFNMSLKRISDRRGAEVTEEEKFTIILFESQSV
GNEILNVQVLTSLFVVIYVHGSQDNNAATKRLMDNAFAEPFVPPVDKLVLPQLC
EALNNKFAEYQNSRGLTKENLVFLAQKLFNNSSHLEDYSLGSLVSWSQFNRENLPGR
NYTFWQFDGVMVELKHLKPHWNDGAILGPFVNKQQAHDLLINKPDGTFLLRFSDEI

GGTIAWKFPDSQERMFNLMPTTRDFSIIRSLADRLGLNLYLYVFPDRPKDEVYSKY
YTPVPCSSATAKAVDGVVKVQIKOVVPEFNASADAGGGSATYMDQAPSVPACQAHY
NMYPQNPDSVLDTDGDLEDTMDVARRVEELLGRPMDSOWIPIHAQS"

ORIGIN

Query Match 99.7%; Score 2356; DB 9; Length 2716;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2359; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ATGGCTGTGTGGATACAGCTCAGCAGCTCAAGGAGAACCCCTTCATCAGATGCAAGCG 60
Db 12 ATGGCTGTGTGGATACAGCTCAGCAGCTCAAGGAGAACCCCTTCATCAGATGCAAGCG 71
Qy 61 TTATATGCCAGCAGCATTTTCCATTGAGTGGCGGCAATTTATTTATCCAGTGGATGAAAGC 120
Db 72 TTATATGCCAGCAGCATTTTCCATTGAGTGGCGGCAATTTATTTATCCAGTGGATGAAAGC 131
Qy 121 CAAGCATGGGACTCAGTAGATCTTTGATAAATCCACAGGAGAAATTAAGGCCACCCAGCTC 180
Db 132 CAAGCATGGGACTCAGTAGATCTTTGATAAATCCACAGGAGAAATTAAGGCCACCCAGCTC 191
Qy 181 CTGGAGGGCTGTGTGGAGGCTGCAGAGGCTGCAGAGAGGAGCAGCAGGTTGGGGAGAGTCGG 240
Db 192 CTGGAGGGCTGTGTGGAGGCTGCAGAGGCTGCAGAGAGGAGCAGCAGGTTGGGGAGAGTCGG 251
Qy 241 TTTTCTCTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACACCTATGACCGC 300
Db 252 TTTTCTCTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACACCTATGACCGC 311
Qy 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCATATATTTGTAACAATGAACAGAGTTGGTC 360
Db 312 TGCCCCATGGAGCTGGTCCGCTGCATCCGCATATATTTGTAACAATGAACAGAGTTGGTC 371
Qy 361 CGAGAAGCCAAATGTAGTCTCTCAGCTGGAGGCTTGTGTGATGCCATCTCCAGAAA 420
Db 372 CGAGAAGCCAAATGTAGTCTCTCAGCTGGAGGCTTGTGTGATGCCATCTCTCCAGAAA 431
Qy 421 CACCTCCAGATCAACACAGACGTTTGGAGGCTGGAGCTGGTGCAGCAGGACACAGAGAAAT 480
Db 432 CACCTCCAGATCAACACAGACGTTTGGAGGCTGGAGCTGGTGCAGCAGGACACAGAGAAAT 491
Qy 481 GAGTAAATAAGCTGCAGCAGACTCAGAGTACTTTATCATCTATCCAGTACAGAGAGCCCTG 540
Db 492 GAGTAAATAAGCTGCAGCAGACTCAGAGTACTTTATCATCTATCCAGTACAGAGAGCCCTG 551
Qy 541 AGGATCCAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGGCTCTGAGCCGG 600
Db 552 AGGATCCAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGGCTCTGAGCCGG 611
Qy 601 GAGACGGCCCTCCAGCAGAGACGAGTGTCTCTGGAGGCTGGTGTGACAGCTGAGGACACAG 660
Db 612 GAGACGGCCCTCCAGCAGAGACGAGTGTCTCTGGAGGCTGGTGTGACAGCTGAGGACACAG 671
Qy 661 ACACCTGACAGTACCGCGTGGAGCTGCCGAGAAAGACACAGAGACCCCTGACGTGCTG 720
Db 672 ACACCTGACAGTACCGCGTGGAGCTGCCGAGAAAGACACAGAGACCCCTGACGTGCTG 731
Qy 721 CGGAAGCAGCAGACCATCATCTCTGGATGACAGCTGATCCAGTGGAGGCGGCGGACGAG 780
Db 732 CGGAAGCAGCAGACCATCATCTCTGGATGACAGCTGATCCAGTGGAGGCGGCGGACGAG 791
Qy 781 CTGGCCGGGAACGGCGGCCCCCGGAGGAGCTGGACGCTGTACAGTCTCTGGTGTGAG 840
Db 792 CTGGCCGGGAACGGCGGCCCCCGGAGGAGCTGGACGCTGTACAGTCTCTGGTGTGAG 851
Qy 841 AAGTTGGCGGAGATCATCTTGGCAGAACCGGACAGAGATCCCGAGGGCTGAGACACTCTGC 900
Db 852 AAGTTGGCGGAGATCATCTTGGCAGAACCGGACAGAGATCCCGAGGGCTGAGACACTCTGC 911
Qy 901 CAGCAGCTGCCCATCCCGGCGCCAGTGAGAGATGCTGGCCGAGGCTCAACGCCACCATC 960
Db 912 CAGCAGCTGCCCATCCCGGCGCCAGTGAGAGATGCTGGCCGAGGCTCAACGCCACCATC 971

QY 961 ACGGACATTATCTCAGCCCTGGTGACAGCAGCTTTCATCATTTGAGAGACGACCTCCTCCAG 1020
Db 972 ACGGACATTATCTCAGCCCTGGTGACAGCAGCTTTCATCATTTGAGAGACGACCTCCTCCAG 1031
QY 1021 GTCTGAAGACCCAGACCAAGTTTTCAGCCACTGTGGCCCTGCTGCTGGGGGGGGAAGCTG 1080
Db 1032 GTCTGAAGACCCAGACCAAGTTTTCAGCCACTGTGGCCCTGCTGCTGGGGGGGGAAGCTG 1091
QY 1081 AACGTGCATGAACACCCCCCGAGGTGAAGGCCACCAATCATCATGTGAGCAGCAGGCCAAG 1140
Db 1092 AACGTGCATGAACACCCCCCGAGGTGAAGGCCACCAATCATCATGTGAGCAGCAGGCCAAG 1151
QY 1141 TCTCTGCTCAAGAAGCAGAACACCGCAATCATTTACAGTGGCGAGATCTTGAACAACTGC 1200
Db 1152 TCTCTGCTCAAGAAGCAGAACACCGCAATCATTTACAGTGGCGAGATCTTGAACAACTGC 1211
QY 1201 TGCCTCATGTGAGTACCAACAGCCACAGGCAACCTTTAGTGGCCACTTTCAGGAATATGTCC 1260
Db 1212 TGCCTCATGTGAGTACCAACAGCCACAGGCAACCTTTAGTGGCCACTTTCAGGAATATGTCC 1271
QY 1261 CTGAAAAGCAATTAAGAGTTCAGACCGTGTGGGGCAGAGTCGGTGACGAAGAAAAATTT 1320
Db 1272 CTGAAAAGCAATTAAGAGTTCAGACCGTGTGGGGCAGAGTCGGTGACGAAGAAAAATTT 1331
QY 1321 ACAATCTCTGTTTGAATCCAGTTCAGTGTGTGCGTGAATGAGCTGTTTTCAGAGTCAAG 1380
Db 1332 ACAATCTCTGTTTGAATCCAGTTCAGTGTGTGCGTGAATGAGCTGTTTTCAGAGTCAAG 1391
QY 1381 ACCCTGTCCCTGCCAGCTGGTGTGATCGTTTCATGGCAGCAGGACCAATTCGCACGGCC 1440
Db 1392 ACCCTGTCCCTGCCAGCTGGTGTGATCGTTTCATGGCAGCAGGACCAATTCGCACGGCC 1451
QY 1441 ACTGTTCTCTGGGACAATGCTTTTTCAGAGCCTGGCAGGAGTGCATTTGCGTGCCTGAC 1500
Db 1452 ACTGTTCTCTGGGACAATGCTTTTTCAGAGCCTGGCAGGAGTGCATTTGCGTGCCTGAC 1511
QY 1501 AAAGTGTCTGGCCACAGCTGTGTGAGCGCTCAACATGAATTCAGAGCCGGAAGTGCAG 1560
Db 1512 AAAGTGTCTGGCCACAGCTGTGTGAGCGCTCAACATGAATTCAGAGCCGGAAGTGCAG 1571
QY 1561 AGCAACCGGGCCCTGACCAAGGAGAACTCGTGTTCCTGGCGCAGAACTGTTCAACAAC 1620
Db 1572 AGCAACCGGGCCCTGACCAAGGAGAACTCGTGTTCCTGGCGCAGAACTGTTCAACAAC 1631
QY 1621 AGCAGCAGCCACCTGGAGGACTACAGTGGCCTGTCTGTCTGCTGGTCCAGTTCACAGG 1680
Db 1632 AGCAGCAGCCACCTGGAGGACTACAGTGGCCTGTCTGTCTGCTGGTCCAGTTCACAGG 1691
QY 1681 GAGAAATTTACCAGACGGAATTTACACTTCTGCGCAATGGTTTGAACGGTGTGATGGAAGTG 1740
Db 1692 GAGAAATTTACCAGACGGAATTTACACTTCTGCGCAATGGTTTGAACGGTGTGATGGAAGTG 1751
QY 1741 TTAATAAACAATCTCAAGCCTCATTTGGAATGATGGGCGCAATTTTGGGGTTTGTAAACAAG 1800
Db 1752 TTAATAAACAATCTCAAGCCTCATTTGGAATGATGGGCGCAATTTTGGGGTTTGTAAACAAG 1811
QY 1801 CAACAGGCCCATGACTACTGATTAACAGCCAGATGGGACCTTCCTCCTGAGATTCAGT 1860
Db 1812 CAACAGGCCCATGACTACTGATTAACAGCCAGATGGGACCTTCCTCCTGAGATTCAGT 1871
QY 1861 GACTCAGAAATTTGGCGGATCACCATTTCTTTGGAAGTTTGAATCTCAGGAAGAAATGTTT 1920
Db 1872 GACTCAGAAATTTGGCGGATCACCATTTCTTTGGAAGTTTGAATCTCAGGAAGAAATGTTT 1931
QY 1921 TGGAAATCTGATGCTCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCCGCGCTTG 1980
Db 1932 TGGAAATCTGATGCTCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCCGCGCTTG 1991
QY 1981 GGAGACTTGAATTAACCTTATCTAGCTGTTTCTGATCGGCCAAAAGATGAAGTATCTCC 2040
Db 1992 GGAGACTTGAATTAACCTTATCTAGCTGTTTCTGATCGGCCAAAAGATGAAGTATCTCC 2051
QY 2041 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2100

Db 2052 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTGTGATGGAATACGTG 2111
QY 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGTAACGCATCTGCAGATGCCGGGGC 2160
Db 2112 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGTAACGCATCTGCAGATGCCGGGGC 2171
QY 2161 GGCAGCCGACGATACATGGAACGAGGCCCTCCCTCCAGCTGTGTGTCGCCAGGCTCATTAT 2220
Db 2172 GGCAGCCGACGATACATGGAACGAGGCCCTCCCTCCAGCTGTGTGTCGCCAGGCTCATTAT 2231
QY 2221 AACATGTACCACACAGAAACCTGACTGACTGCTTGTGACACCGATGGGAGCTTCGATCTGGAG 2280
Db 2232 AACATGTACCACACAGAAACCTGACTGACTGCTTGTGACACCGATGGGAGCTTCGATCTGGAG 2291
QY 2281 GACAAATGGAAGTAGCGCGCGCTGTGGAGAGCTCCTGGGCGGCAATGGACAGTCAAG 2340
Db 2292 GACAAATGGAAGTAGCGCGCGCTGTGGAGAGCTCCTGGGCGGCAATGGACAGTCAAG 2351
QY 2341 TGGATCCGCGACGACCAATCGTGA 2364
Db 2352 TGGATCCGCGACGACCAATCGTGA 2375

RESULT 9

BC065227 LOCUS 5167 bp mRNA linear PRI 12-OCT-2004
DEFINITION Homo sapiens cDNA clone MGC:74606 IMAGE:5553426, complete cds.
ACCESSION BC065227

VERSION BC065227.1 GI:40807019

KEYWORDS MGC

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5167)

REFERENCE

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Adams, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.W.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

JOURNAL

PUBMED 2 (bases 1 to 5167)

REFERENCE

Director MGC Project.

AUTHORS

Direct Submission

TITLE

Submitted (12-JAN-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: gcapsb@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

Genome Sequencing: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;
 web site: <http://www.nisec.nih.gov/>
 Contact: nisc_mcg@nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masliello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/Link at: <http://image.llnl.gov>

Series: IRAP Plate: 139 Row: b Column: 4
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 42519913.

FEATURES

source

Location/Qualifiers
 1..5167
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:74606 IMAGE:5553426"
 /tissue type="Lymph, lymphoma"
 /clone_lib="NIH MGC_85"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 170..2533

CDS

/codon_start=1
 /product="Unknown (protein for MGC:74606)"
 /protein_id="AAH65227.1"
 /db_xref="GI:40807020"
 SVDLNDPOENTKATQLLEGLVQLQKAEHVGEDGFLKIKLGHATQLQNTDRCP
 MELVRCIRHILYNLRVREANNSSPAGSLADAMQKHLQINOTFSELRLVTDQEN
 ELKLOQYEFYIYOYSLRIQAFQPLAQSPQLRSLRQALQKQVSLQVLEAWLQRE
 AQTLLQYVELAEKHQKTLQLLRKQQTIIIDDELIQWRKRLQAGNPPGSLDLVQ
 SWCKPLAIWNRQOIRRAEHLQQLPIPGPVEMLAENVNATIDIIISALVTSTFI
 KQPPVLTQTKTFAATVRLVGLKLVNHPVQKATIISEQAKSLIKNENTNDY
 SGEILNVCVMEYHQATGLSAHFRNLSLKKIKRDRGASVTEETFTILFESQFSV
 GKNELVQVLTSLPVLVYVHGSQDNNTATVLDNAPAEPRVFPVAVDKVLPQLC
 EALNKQFAEYVSRGTLKENVLAQLKFNSSSHLEDYSLGSVQFNRLNLPGR
 NYTFQWFDGVEMLKXHLKPHNDGAILGSRVQKQADHLINKPDGTFLILRFDSEI
 GGITIAKFDQSEMFNMLPMTDFRSIRSLADRLGLDNLVLIYVFPDRPKDEVSKY
 YTPVPCSATAKADVGVKPKQVPEFVNASADGGSATYMDQAPSPAVCPQAHY
 NMYFQNPUSVLDTDFDLEDMVARRVBEELRPMDSQWIPHAQS"

ORIGIN

Query Match 99.7%; Score 2356; DB 9; Length 5167;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2359; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 1 ATGGCTGTGGATCAAGCTCAGAGCTCAAGGAGAGAGGCTTTCATCAGATGCAAGCG 60
 170 ATGGCTGTGGATCAAGCTCAGAGCTCAAGGAGAGAGGCTTTCATCAGATGCAAGCG 229
 61 TTATATGCCAGCATTTTCCATTGAGCTGGGCTATTTATCCAGTGGATGAAAGC 120
 230 TTATATGCCAGCATTTTCCATTGAGCTGGGCTATTTATCCAGTGGATGAAAGC 289
 121 CAAGCATGGGACTCAGTAGACTTCTATATCCACAGGAGAACATTAAGGCCACCCAGCTC 180
 290 CAAGCATGGGACTCAGTAGACTTCTATATCCACAGGAGAACATTAAGGCCACCCAGCTC 349
 181 CTGAGGGCCCTGGTCAGGAGCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
 350 CTGAGGGCCCTGGTCAGGAGCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 409
 241 TTTTACTGAAGATCAAGCTGGGGCACTATGCGCACAGCTCCAGAACACGATGACCGC 300
 410 TTTTACTGAAGATCAAGCTGGGGCACTATGCGCACAGCTCCAGAACACGATGACCGC 469
 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCCATATATTGTACAATGAACAGAGGTTGCTC 360

Db Qy 470 TGCCCCATGGAGCTGGTCCGCTGCATCGCCATATATTGTACAATGAACAGAGGTTGCTC 529
 Qy 361 CGAAGGCCAACATGCTAGCTCTCCAGCTGGAGCCTTGTGATGCTCCAGGAA 420
 Db 530 CGAAGGCCAACATGCTAGCTCTCCAGCTGGAGCCTTGTGATGCTCCAGGAA 589
 Qy 421 CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGCAGCTGGTCCAGCAGGACACAGAGAT 480
 Db 590 CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGCAGCTGGTCCAGCAGGACACAGAGAT 649
 Qy 481 GAGTTAAAAAGCTGCAGCAGACTCAGAGTACTTTCATTCATCCAGTATCCAGGAGCCTG 540
 Db 650 GAGTTAAAAAGCTGCAGCAGACTCAGAGTACTTTCATTCATCCAGTATCCAGGAGCCTG 709
 Qy 541 AGGATCCAAGCTCAGTTTGGCCCTGCTGCCAGCTGAGCCCTCCAGGAGCCTTCCAGCCGG 600
 Db 710 AGGATCCAAGCTCAGTTTGGCCCTGCTGCCAGCTGAGCCCTCCAGGAGCCTTCCAGCCGG 769
 Qy 601 GAGAGCGCCCTCCAGCAGAGCAGGTGTCTCTGGAGGCTCTGCTGAGCGTGGAGCAG 660
 Db 770 GAGAGCGCCCTCCAGCAGAGCAGGTGTCTCTGGAGGCTCTGCTGAGCGTGGAGCAG 829
 Qy 661 ACCTGTCAGCAGTACCGCTGAGCTGCCCGAGAGGACACAGAGAGCCTTCCAGGCTGCTG 720
 Db 830 ACCTGTCAGCAGTACCGCTGAGCTGCCCGAGAGGACACAGAGAGCCTTCCAGGCTGCTG 889
 Qy 721 CGAAGCAGCAGCAGCATCTCTGATGACAGCTGATCCAGTGGAGGCGGCGGAGCAG 780
 Db 890 CGAAGCAGCAGCAGCATCTCTGATGACAGCTGATCCAGTGGAGGCGGCGGAGCAG 949
 Qy 781 CTGCGCGGGAACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
 Db 950 CTGCGCGGGAACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1009
 Qy 841 AAGTTGGCGGAGATCATCTTGGCAGAAACGGCAGCAGATCCGAGGCTTGAGCAGCTCTGTC 900
 Db 1010 AAGTTGGCGGAGATCATCTTGGCAGAAACGGCAGCAGATCCGAGGCTTGAGCAGCTCTGTC 1069
 Qy 901 CAGCAGCTGCCATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 Db 1070 CAGCAGCTGCCATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1129
 Qy 961 ACGCACATTAATCTCAGCCCTGGTCCAGCAGCAGCTTTCATTCATTTGAGAAGCAGCCTCTCAG 1020
 Db 1130 ACGCACATTAATCTCAGCCCTGGTCCAGCAGCAGCTTTCATTCATTTGAGAAGCAGCCTCTCAG 1189
 Qy 1021 GTCTGAAGACCCAGACCAAGTTTTCAGCCACTGTGCGCCTGTGCGCTGTGCGGCGGGAAGCTG 1080
 Db 1190 GTCTGAAGACCCAGACCAAGTTTTCAGCCACTGTGCGCCTGTGCGCTGTGCGGCGGGAAGCTG 1249
 Qy 1081 AACGTGCACATGAACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
 Db 1250 AACGTGCACATGAACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1309
 Qy 1141 TCTCTGTCTCAAGAACAGAGAACACCGCAATGATTAAGTGGCGAGATCTTGAACAACTGC 1200
 Db 1310 TCTCTGTCTCAAGAACAGAGAACACCGCAATGATTAAGTGGCGAGATCTTGAACAACTGC 1369
 Qy 1201 TGCGTCAATGAGTACCAAGCAGGCAACCGCACTTGTAGTGGCCCTTCCAGGAATATGTC 1260
 Db 1370 TGCGTCAATGAGTACCAAGCAGGCAACCGCACTTGTAGTGGCCCTTCCAGGAATATGTC 1429
 Qy 1261 CTGAAAACGAATTAAGAGTACAGCGCTGTGGGCGAGAGTGGTGCACAGAGAGAGAGAGAGAG 1320
 Db 1430 CTGAAAACGAATTAAGAGTACAGCGCTGTGGGCGAGAGTGGTGCACAGAGAGAGAGAGAGAG 1489
 Qy 1321 ACAATCTGTTGATCCAGTTCAGTGTTCGTAATAGCTGTTTTCAGTCAAGTCAAG 1380
 Db 1490 ACAATCTGTTGATCCAGTTCAGTGTTCGTAATAGCTGTTTTCAGTCAAGTCAAG 1549
 Qy 1381 ACCCTGTCCCTGCCAGTGGTGGTGTGATCGTTTCATGGCAGCAGGACCAACATGCGACGCGCC 1440


```

Qy 2341 TGGATCCGCGACACAACTCGTGA 2364 Db
|||||
Db 2353 TGGATCCCTCACGCGCAGTGTGA 2376 Qy
|||||

RESULT 11
BTA237933 BTA237933 2485 bp mRNA linear MAM 15-JUN-2000
LOCUS Bos taurus mRNA for mammary gland factor STAT5B.
DEFINITION AJ237933
VERSION AJ237933.1 GI:5689666
KEYWORDS Mammary Gland Factor STAT5B; stat5b gene.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1
REFERENCE
AUTHORS Seyfert,H.M., Pitra,C., Meyer,L., Brunner,R.M., Wheeler,T.T.,
Molenaar,A., McCracken,J.Y., Herrmann,J., Thiesen,H.J. and
Schwerin,M.
TITLE Molecular characterization of STAT5A- and STAT5B-encoding genes
reveals extended intragenic sequence homogeneity in cattle and
mouse and different degrees of divergent evolution of various
domains
JOURNAL J. Mol. Evol. 50 (6), 550-561 (2000)
MEDLINE 20296816
PUBMED 10835485
REFERENCE 2 (bases 1 to 2485)
AUTHORS Seyfert,H.M.
TITLE Direct Submission
SUBMITTED (30-MAR-1999) Seyfert H.M., Molecular Biology, Research
Institute for the Biology of Farm Animals, Wilhelm-Stahl-Allee-2,
18196 Dummerstorf, GERMANY
FEATURES
source
1..2485
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="mammary gland"
/dev_stage="lactating"
1..2485
/gene="stat5b"
2..2365
/gene="stat5b"
/codon_start=1
/product="mammary gland factor STAT5B"
/protein_id="CAB51845.1"
/db_xref="GI:5689667"
/db_xref="GOA:Q9TUM3"
/db_xref="UniProt/Swiss-Prot:Q9TUM3"
/translation="MAVWIAQQLQGDALHQMQLYQHPPIEVRYHLSQWIEQAWD
SIDLNPQENIKATLQELGVOLKKAHQVGEDGFLKXIKLGHVATQLQNTYDRCP
MELVRCIRHILYNEORLAVREANNQSPAGSFADAMSQKHQIQINQTEELRLVTDREN
ELKLLQTOBYFIIQYQESIRIQAPQALQALPOERLSRETALQKQVSLKAWLQRE
AQTLQTRVLEAKHQKTLQRLKQQTIIIDDELIQWRKQQLAGNGPPEGLDVLQ
SWCKEALAIWQNRQIIRRAHLCCQLPIGPVEMLAEVYNATIDIIISALVTSFII
EKOPPVLYKTQTFRAVIRLLVGGKLVNHNPPQVKATIISEQAKSLKKNENTRNDY
SGBILNCCVMYHQATGTLSAHFRNMSLKRIRKSDRRGAESVTEKFTILFESQSV
GNELVFOVKTSLPVVVI VHGSQDNNTATVLWDNAFAPPGVPAVPDKVLWPOLC
BALNMKEKAEVQSNRGLTKENLVFLAQKLFNSSSHLEDYNGMSVSKSQPNRELPR
NYTFQWDFQSGEVLKHLKPHWNDGAILGFVNRKQHLINKSDGTFLLRFSDEI
GGTIAWQDSQERMFNWLFFTRDPSIRSLADRLGDSLYLYVPPDRPKRVYSKY
YTFVPCPATAKVDGVYKPKQIKVVPVEFVSASADSAGGSATYMDQAPSVAVCPQPHY
NMYPQNPDPVLDNDGDPDLDDTIDVARRVEELLGRPMDSQWIPHAQS"

Query Match 88.0%; Score 2079.2; DB 4; Length 2485;
Best Local Similarity 92.5%; Pred. No. 0;
Matches 2186; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Qy 1 ATGCGCTGTGGATACAGCTCAGCAGCTCCAGAGAACGCCCTTCATCAGATGCAAGCG 60
|||||
2 ATGCGCTGTGGATACAGCTCAGCAGCTCCAGAGGCGATGCCCTTCATCAGATGCAAGCG 61
|||||
61 TTATATGGCCAGCATTTTCCCATTTAGGCTCGGCATTTATTTATCCAGTGGATGAAGC 120
|||||
62 CTGTACGGCCAGCATTTCCCATTTAGGCTCGGCATTTACTTTATCCAGTGGATGAAGCG 121
|||||
121 CAAGCATGGAGCTCAGTAGATCTTGATATCCACAGAGAACCATTTAAGGCCACCCAGCTC 180
|||||
122 CAAGCTTTGGGACTCAATAGATCTTGATATCCACAGAGAACCATTTAAGGCCACCCAGCTC 181
|||||
181 CTGAGGGCCCTGTGTCAGGAGCTGCAGAGAAGCGAGCACAGGTGGGGGAAGATGGG 240
|||||
182 CTGAGGGCCCTGTGTCAGGAGTTGCGAGAAGCGAGCACCAAGTGGGGGAAGACGGG 241
|||||
241 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC 300
|||||
242 TTCTTACTGAAGATCAAGCTGGGGCACTACGCCACACAGCTCCAGAACACATATGACCGC 301
|||||
301 TGCCCATGGAGCTGGTCCGCTGCATCCGCGCATATATTGTACATGNAAGAGGTTGGTC 360
|||||
302 TGCCCATGGAGCTGGTCCGCTGCATCCGCGCATATTTCTGTACATGNAAGAGGTTGGTC 361
|||||
361 CGAGAAGCCCAACATGCTAGCTCTCCAGCTGGAGGCTTCTGCTGATGCCATGTCCAGAAA 420
|||||
362 CGAGAAGCCCAACATGCTAGCTCTCCGCTGGAGGTTTCCGCGATGCCATGTCCAGAAA 421
|||||
421 CACCTCCAGATCAACAGACGTTTGGAGAGCTCGACTGCTGCACGAGGACACAGAGAA 480
|||||
422 CACCTTCAATCAACAGACATTTGAGGAGCTCGACTGCTGCACAGAACACAGAGAA 481
|||||
481 GAGTTAAAAGCTGCAGCAGACTCAGAGTACTTTCATCATCAGATCAGAGGAGGCTG 540
|||||
482 GAGCTGAAGAAGCTGCAGCAGACTCAAGAGTACTTTTATCATCCAGTATCAGAGGAGCTG 541
|||||
541 AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGAGAGCGCTCTGAGCCG 600
|||||
542 AGGATCCAGGCTCAGTTTGGCCAGCTGGCCAGCTGNAACCCCGAGAGCGACTGAGCCG 601
|||||
601 GAGACGGCCCTCCAGCAGAACAGAGGTGTCTCTGGAGGCTGGTTGACGCTGAGGACAG 660
|||||
602 GAGACGGCCCTCCAGCAGAACAGAGGTGTCTCTGGAGGCTGGCTGAGCGCGAGGCCAG 661
|||||
661 ACAGTGCAGCAGTACCGCTGGAGCTGCCCGAGAGCACACAGAACCCCTGCAGCTGCTG 720
|||||
662 ACCTGTCAGCAGTACCGCTGGAGCTGGCCGAGAGTGGCCGAGAGCACACAGAACCCCTGCAGCTGCTG 721
|||||
721 CGGAAGCAGCAGACCATCATCTCTGGATGACGAGCTGATCCAGTGGAGCGCGCGCAGCAG 780
|||||
722 CGGAAGCAGCAGACCATCATCTCTGGATGACGAGCTGATCCAGTGGAGAGCGCGCGCAGCAG 781
|||||
781 CTGGCCGGGAAACCGCGGGCCCCCGAGGCGAGCTGCGACGTGTACAGTCTCTGTGTGAG 840
|||||
782 CTGGCCGGGAAACCGAGGGCCCCCGAGGCGAGCTGCTGCTACAGTCTCTGTGTGAG 841
|||||
841 AAGTTGGCGGAGATCATCTGGCAGAACCGCGCAGACTCCGAGGGCTGAGCAGCTCTGC 900
|||||
842 AAGTTGGCGGAGATCATCTGGCAGAACCGCGCAGACTCCGAGAGCGCGCAGCAGCTCTGC 901
|||||
901 CAGCAGCTGCCATCCCGCGCCAGTGGAGGAGATGCTGGCCGAGGTCGACCGCACCATC 960
|||||
902 CAGCAGCTGCCATCCCGCGCCAGTGGAGGAGATGCTGGCTGAGGTCAACCGCACCATC 961
|||||
961 ACGGACATTTATCTCAGCCCTGGTGA CACAGACGTTTATCATATTGAGAGAGAGCTCTCTCAG 1020
|||||
962 ACTGACATCATCTCAGCCCTGGTGA CACAGACGTTTATCATCATCAGAGAGAGAGCTCTCAG 1021
|||||
1021 GTCTGAGAGCCAGACCAAGTTTGCAGCCACTGTGGCGCTGCTGGTGGGGGAGAGCTG 1080
|||||
1022 GTCTGAGAGCCAGACCAAGTTTGGCGGCGCACCGTGGCGCTGCTGGTGGGGGAGAGCTG 1081
|||||
1081 AAGCTGCACATGAACCCCCCGAGGTGAAGGCCACCATCATCATGTCAGAGCAGAGCGGCAAG 1140
|||||
1 ATGCGCTGTGGATACAGCTCAGCAGCTCCAGAGAACGCCCTTCATCAGATGCAAGCG 60

```


Query Match		84.2%;	Score 1991.2;	DB 10;	Length 2671;
Best Local Similarity		90.4%;	Pred. No. 0;		
Matches 2138;		Conservative	0;	Mismatches 223;	Indels 3; Gaps 1;
Qy	1	ATGGCTGTGGATACAAAGCTCAGCAGCTCCAGGAGAACCCCTTCATCAGATCGAAGCG	60		
Db	111	ATGGCAATGTGGATACAGGCTCAGCAGCTCAGGGCGATGCCCTTCATCAGATCGAGGCC	170		
Qy	61	TTATATGCCAGCAATTTTCCATTGAGGTGGGCAATTTATTTATCCCAAGTGATTTGAAGC	120		
Db	171	TTGTACGGCCAGCAATTTCCCATGCTGAGGTGGCAATTTATTTATCCCAAGTGATCGAAGT	230		
Qy	121	CAAGCATGGGACTCAGTAGATCTTGATTAATCCACAGGAGAACTTAAGGCCACCCAGCTC	180		
Db	231	CAAGCTGGGACTCAATAGATCTTGATTAATCCACAGGAGAACTTAAGGCCACCCAGCTC	290		
Qy	181	CTGAGGGCTGTGTGACAGGAGCTGCAGAGAAGGCAGACACAGGTGGGGAGATGGG	240		
Db	291	CTGAGGGCTGTGTGACAGGAGCTGCAGAGAAGGCAGACACAGGTGGGGAGATGGG	350		
Qy	241	TTTTTACTGAAGATCAAGCTGGGCACTATGCCACAGCTCCAGAAACAGTATGACCGC	300		
Db	351	TTTTTGTGAAGATCAAGCTGGGCACTATGCCACAGCTCCAGAAACAGTATGACCGC	410		
Qy	301	TGCCCATGGAGCTGGTCCGTGATCCGGCATATATTGTACAAATGAACAGAGTTGGTC	360		
Db	411	TGCCCATGGAGCTGGTCCGTGATCCGGCATATTTATACAAATGAACAGAGCTGGTC	470		
Qy	361	CGAGAAGCAACATGTAGCTTCCAGCTGGAGCCCTGCTGATGCTCATGTCCAGAAA	420		
Db	471	CGAGAAGCAACATGTAGCTTCCAGCTGGAGCCCTGCTGATGCTCATGTCCAGAAA	530		
Qy	421	CACCTCCAGATCAACACAGAGCTTGAGGAGCTGCGACTGTGTCAGCGAGACACAGAGAT	480		
Db	531	CACCTTCAGATCAACCAACAGTTTGAGGAGCTGCTGATCAACACAGGACACGGAGAGC	590		
Qy	481	GAGTTAAAAAGCTGACGACAGACTCAGGAGTACTTTCATCTCATGTACAGGAGAGCTG	540		
Db	591	GAGCTGAAGAGCTGACGACAGACCAAGAGTACTTTCATCTCATGTACAGGAGAGCTA	650		
Qy	541	AGGATCCAGCTCAGTTTGGCCGCTGCCAGCTGAGCCGCCAGGAGCGCTCTGAGCCGG	600		
Db	651	CGGATCCAGCTCAGTTTGGCCGCTGCCAGCTGAGCCGCCAGGAGCGCTCTGAGCCGG	710		
Qy	601	GAGACGGCCCTCCACGAGAGAGAGGTGTCTCTGGAGCCCTGGTTGACGCGTGAAGGCACG	660		
Db	711	GAGACGGCCCTCCACGAGAGAGAGGTGTCTCTGGAGCCCTGGCTGCAGCGAGAGGCACAG	770		
Qy	661	ACACTGACAGATACCGGTGGAGCTGCCGAGAGACACAGAGACCTTGACGCTGCTG	720		
Db	771	ACACTGACAGATACCGGTGGAGCTGCCGAGAGACACAGAGACCTTGACGCTGCTG	830		
Qy	721	CGGAAGCAGCAGACCATCTCTGGATGACGAGCTGATCGAGTGAAGCGGCGCAGCAG	780		
Db	831	AGGAGCAGCAGACCATCTCTGGATGACGAGCTGATCGAGTGAAGCGGCGCAGCAG	890		
Qy	781	CTGGCCGGGAACGGCGGGCCCCCGAGGCGAGCCTGGACGTGTCTACAGTCTCTGGTGTGAG	840		
Db	891	CTGGCCGGGAATGGGGGTCTCTCTGAGGCGAGCCTGGATGTGCTGCAGTCTCTGGTGTGAG	950		
Qy	841	AAGTTGGGGAGATCATCTGCGAGAACCGGCGAGCAGATCCGAGGGCTGAGCCTCTGC	900		
Db	951	AAGCTGCCGAGATCATCTGCGAGAACCGGCGAGCAGATCCGAGGGCTGAGCCTCTGC	1010		
Qy	901	CAGCAGCTGCCATCCCGGGCCAGTGGAGGAGATGCTGGCCGAGGTCAACGCCACCATC	960		
Db	1011	CAGCAGCTGCCATCCCGGGCCAGTGGAGGAGATGCTGGCCGAGGTCAACGCCACCATC	1070		
Qy	961	ACGGACATTTCTCAGCCCTGGTGACCAAGCAGTTCATCTATTTGAGAGACGACCTCTCAG	1020		
Db	1071	ACGACATCTCTCAGCCCTGGTGACCAAGCAGTTCATCTATTTGAGAGACGACCTCTCAG	1130		

Qy	1021	GTCTGAAAGCCACGACCAAGTTTTCAGACCCACTGTGCGCCTGCTGGTGGGGGGAAGCTG	1080
Db	1131	GTCTGAAAGCCACGACCAAGTTTTCAGACCCACTGTGCGCCTGCTGGTGGGGGGAAGCTG	1190
Qy	1081	AAGTGCACATGAACCCGCCAGGTGAAGGCCACCATCATCAGTGAAGCAGCAGGCCAAG	1140
Db	1191	AAGTGCACATGAACCCCTCGAGGTGAAGGCCAACCATCATCAGCAGCAGCAGGCCAAG	1250
Qy	1141	TCCTGCTCAAGAACGAGAACACCCGCAATGATTAACAGTGGCGAGATCTTTGAACAACTGC	1200
Db	1251	TCCTGCTCAAGAACGAGAACACCCGCAATGATTAACAGCAGCAGATCTCTGAACAACTGC	1310
Qy	1201	TGCGTCATGAGATACCAACAGCAGCAGCACCCTTAGTGGCCCACTTCAGAGATATGCC	1260
Db	1311	TGCGTCATGAGATACCAACAGGCGCAGCAGCACCCTCAGCGCCCACTTCAGAGATATGCC	1370
Qy	1261	CTGAACCAATTAAGAGGTGAGACCCGTGCTGGGCGCAGAGTCGGTGAACAGAAATAATTT	1320
Db	1371	CTGAACCAATTAAGAGGTCTGACCCGCTGGTGCAGAGTCGGTCAAGAGAGAGAGTTC	1430
Qy	1321	ACAACTCTGTTGAATCCCAAGTTCAAGTTGTTGGTGGAAATGAGCTGGTTTTCAAGTCAAG	1380
Db	1431	ACAACTCTGTTGACTCACAGTTTCAAGTTGTTGGTGGAAATGAGCTGGTTTTCAAGTCAAG	1490
Qy	1381	ACCTGTCTCTGCGCAGCTGGTGTGATCGTTTCATGGCAGCCAGCAGCAACAATGCGACGGC	1440
Db	1491	ACCTGTCTCTGCGGTGGTGGTGTGATCGTTTCACGGCAGCCAGCAGCAACAATGCGACGGC	1550
Qy	1441	ACTGTTCTCTGGGCAATGCTTTTTCAGAGAGCTTGGCAGGCTGCATTTGCGCTGCCTGAC	1500
Db	1551	ACCGTCTCTGGGCAACGCTTTTCAGAGAGCTTGGCAGGCTGCATTTGCGCTGCCTGAC	1610
Qy	1501	AAAGTGTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAATAATCAAGCGCGAAGTGCAG	1560
Db	1611	AAGTGTGTGTGGCCAGCTGTGTGAGGCGCTCAACATGAATAATCAAGCGCTGCAGTGCAG	1670
Qy	1561	AGCAACGGGGCTGACCAAGGAGAACCTGCTGCTCTGGCGCAGAGAACCTGTTCAACAAC	1620
Db	1671	AGCAACGGGGCTGACCAAGGAGAACCTGCTGCTCTGGCGCAGAGAACCTGTTCAACAGC	1730
Qy	1621	AGCAGCAGCCACCTGGAGGACTACAGTGGCTGTCTGTGTCTCTGGTCCCAAGTTCAACAGG	1680
Db	1731	AGCAGCAACACCTCGAGGACTACACAGCATGTTCGTGTCTGGTCCCAAGTTCAACCGG	1790
Qy	1681	GAGAAATTAACAGGACGGAATTAACATTTCTGGCAATGGTTTGAACGCTGTGAAGAGTG	1740
Db	1791	GAGAAATTCGACGAGCGGAATTAACATTTCTGCGAGTGGTTTGAACGCTGTGAAGAGTG	1850
Qy	1741	TTAAAGAACATCTCAGCCTCATTTGGAATGATGGGSCCATTTTGGGTTTTGTAACAACAG	1800
Db	1851	TTAAAGAACATCTCAGCCTCATTTGGAACGATGGGSCCATTTGGGTTTTGTAACAACAG	1910
Qy	1801	CAACAGGCCCATGACCTTACTGATTAAACAGCCAGATGGGACCTTCTCTCTGAGATTTCAGT	1860
Db	1911	CAGAGGCCCATGACCTTACTGATTAAACAGCCAGATGGGACCTTCTCTCTGAGATTTCAGT	1970
Qy	1861	GACTCAGAAATTCGCGGCATCATTGCTTGGAGTTTGAATTTCTCAGGAAAGAAATGTTT	1920
Db	1971	GACTCAGAAATTCGCGGCATCATTGCTTGGAGTTTGAATTTGACTCCCAAGAGAGAAATGTTT	2030
Qy	1921	TGGAATCTGATGCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCGCCGCTTG	1980
Db	2031	TGGAATCTGATGCTTTTACCACAGAGACTTCTCTATCCGTTCCCTCGCTGACCCGCTTG	2090
Qy	1981	GGAGACTTGAATTAACCTTACTACGTGTTTTCTGATCGGCCAAAAGATGAAGTATATCTCC	2040
Db	2091	GGGGAACCTGAATTAACCTTACTACGTGTTTTCTGATCGGCCAAAAGATGAAGTATATCT	2150
Qy	2041	AAATATACACACAGTTTCCCTCGGAGTCTGCTACTGCTAAAGCTGTGTGATGATAGTG	2100
Db	2151	AAGTACTACACACAGTCCCGTGTGAGCCGCCCAACCGGAAAGCAGCTGACGATACGTG	2210
Qy	2101	AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCACTCTGCAGATGCCGGGGC	2160

[illegible]

RESULT 13	BC024319	4893 bp	linear	ROD 06-OCT-2003
LOCUS	BC024319			
DEFINITION	Mus musculus signal transducer and activator of transcription 5B, mRNA (cDNA clone MGC:18421 IMAGE:4237132), complete cds.			
ACCESSION	BC024319			
VERSION	BC024319.1	GI:191354220		
KEYWORDS	MGC.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			

Dickson, M., Schmutz, J., Grimwood, J.; Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 25 Row: e Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7242208.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10

```

1..4893
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:18421 IMAGE:4237132"
/tissue_type="kidney, normal, 5 month old male mouse."
/clone_lib="NCI CGAP_Kid14"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..4893
/gene="Stat5b"
/db_xref="LocusID:20851"
/db_xref="MGI:103035"

```

CDS

CDS

misc feature

/gene="STAT3"; Region: STAT protein, all-alpha domain. STAT
 /note="STAT3; Proteins (Signal Transducers and Activators of
 Transcription) are a family of transcription factors that
 are specifically activated to regulate gene transcription
 when cells encounter cytokines and growth factors. STAT
 proteins also include an SH2 domain pfam00017"
 /db_xref="CDD:pfam01017"
 1147..1902

misc feature

/gene="Stat5b"
 /note="STAT bind; Region: STAT protein, DNA binding domain. STAT proteins (Signal Transducers and Activators of Transcription) are a family of transcription factors that are specifically activated to regulate gene transcription when cells encounter cytokines and growth factors. This family represents the DNA binding domain of STAT, which has an ig-like fold. STAT proteins also include an SH2 domain pfam00017"
 /db_xref="CDD:pfam02854"
 1939..2163

misc feature

```

1994: 12:00
       /gene="Stat5b"
       /note="SH2; Region: SH2 domain"
       /db xref="CPD:pfam00017"

```

ORIGIN

Query Match 83.8%; Score 1981.6; DB 10; Length 4893;
Best Local Similarity 90.2%; Pred. No. 0;

REMARK COMMENT	<p>USA</p> <p>NIH-MGC Project URL: http://mgc.nci.nih.gov</p> <p>Contact: MGC help desk</p> <p>Email: cgapbs@mail.nih.gov</p> <p>Tissue Procurement: Jeffrey E. Green, M.D.</p> <p>CDNA Library Preparation: Life Technologies</p> <p>CDNA Library Arrayed by: The I.M.A.G.E. Center</p> <p>DNA Sequencing by: Sequencing Group at the Sanger Center, Stanford University School of Medicine</p> <p>Web site: http://www-shgc.stanford.edu</p> <p>Contact: (Dickson, Mark) mcd@mail.stanford.edu</p>
-------------------	--

Qy 2161 GGACGGCCAGTACATGACGACGACGCCCCCTCCAGCTGTGTGTCGCCAGGCTCACTAT 2220
Db 2314 G---GGCCACCTACATGATCAGGCTCTCTCCCGAGTCGTGTGCTCCCTCAGGCTCACTAC 2370
Qy 2221 AACATGTACCCACAGAACCTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Db 2371 AACATGTACCCACCAACCCGAGTCCGCTCTTGATACCGATGGGACTTCGATCTGGAA 2430
Qy 2281 GACACATGGACGTAGCGCGCTGTGGAGAGCTCTTGGCGCCGCAATGGACAGTCAG 2340
Db 2431 GACACGATGGACGTGGCGCGCGGTGGAGAGCTCTTAGCGCGCCCATGGACAGTCAG 2490
Qy 2341 TGGATCCCGCAGCACCAATCTGTA 2364
Db 2491 TGGATCCCTCAGCACAGTCATGA 2514

RESULT 14

AY040231

LOCUS 2940 bp mRNA linear ROD 15-MAR-2004
DEFINITION Mus musculus signal transducer and activator of transcription 5B
(Stat5b) mRNA, complete cds.

ACCESSION AY040231

VERSION AY040231.1 GI:21654824

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2940)
Davodi-Semiomi, A., Laloraya, M., Kumar, G.P., Purohit, S., Jha, R.K.
and She, J.-X.

A Mutant Stat5b with Weaker DNA Binding Affinity Defines a Key

Defective Pathway in Nonobese Diabetic Mice

J. Biol. Chem. 279 (12), 11553-11561 (2004)

14701862

REFERENCE 2 (bases 1 to 2940)

Davodi-Semiomi, A., Laloraya, M., Kumar, P.G. and She, J.-X.

Direct Submission

Submitted (13-JUN-2001) Pathology, University of Florida,

Gainesville, FL 32610, USA

Location/Qualifiers

1..2940

FEATURES

source

gene

CDS

/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="caxon:10090"
1..2940
/chromosome="11"
/gene="Stat5b"
484..2844
/gene="Stat5b"
/codon_start=1
/product="signal transducer and activator of transcription
5b"

/protein_id="AAK74074.1"
/db_xref="GI:21654825"
/translation="MAMWIAQQLQGDALHOMQALYGOHPFIEVRHVLISOWIESQAND
SIDLDPENIKATLLEGLVQELQKAEHQVGEDFLIKIKLGHYATLQSTYDRCP
MELVCRTHILYNORLVREANGSPAGSLADAMSKHLQINQTFEELRLITQDTN
ELKLIQTOEYFIYQESLRIOAQGLQNPQERMSRETALQOKQSVLEWLQRE
ACTLQOYVEELAEKHOKTLQLLRKQOTIILDELIQWRKROLAGNGPPPEGLDVLQ
SNCKLAEIWNQROIPRAEHLCOQLPGPVEMLAEVNAITDIIISAMVTFTII
EKQPPVLKTQTKFATVRLVLVGLKLVNMPPOVKATIISEQQAKSLKNNENTNDY
SGEILNCCVMEYHQTATLSAHFNMSLTKRISDRRGAESVTEEFILFDQFSV
GNEILVQVLTSLSPVVIIVHGSQDNATATVMDNAFAEPCRPVFPVDPDLWPLQC
BALNNKFAEYNGRLKENLVFLAQKLFNISNHLNEDYNSMSVSNQFNRENLPGR
NYTFQWFDGVMVEVKHLKPHWMDGAILGFVNQQAHDLLINKPDGTFLLRFSDSI
GGITIAWQDSQERFMNLPETTFDSIRSLADRLGLNLYLVVFPDRPKDEVYSKY
YTPVCEPATKAADGYSKPKQVVPFANASTDAGSATYMDQAPSVPVQPAHYN
MYNPPDSVLATDGLDLEDTDVMDVARRVEELLGRPMDSQWIPHAQS"

ORIGIN

Query Match 83.8%; Score 1980; DB 10; Length 2940;
Best Local Similarity 90.1%; Pred. No. 0;
Matches 2131; Conservative 0; Mismatches 230; Indels 3; Gaps 1;
Qy 1 ATGGCTGTGTGGATACAGCTCAGCAGCTCCAGGAGAGCCCTTCATCAGATCAAGCG 60
Db 484 ATGGCTATGTGGATACAGGCTCAGCAGCTCAGGGCGATGCCCTTCCAGATGAGGCC 543
Qy 61 TTATATGCCAGCAGCATTTTCCCATTTGAGTGGCGCATTTATTTATCCAGTGGATTGAAGC 120
Db 544 TTGTACGGCCAGCATTTCCCATTCAGAGTGGCATTTATTTATCAGATGGATCGAAGC 603
Qy 121 CAGCATGGGACTCAGTAGATCTTTGATAATCCACAGAGAACATTAAGGCCACCCAGCTC 180
Db 604 CAAGCTTGGGACTCAATAGATCTTTGATAATCCACAGAGAACATTAAGGCCACCCAGCTC 663
Qy 181 CTGGAGGCGCTGTGTCAGGAGCTGCAGAGAGCAGACACAGAGTGGGGAAGATGGG 240
Db 664 CTGGAGGCGCTGTGTCAGGAGCTGCAGAGAGCGGAGACACAGTGGGGAAGATGGG 723
Qy 241 TTTTCTGTAAGATCAAGCTGGGCACTATGCCACACAGCTCCAGAAACAGTATGACCGC 300
Db 724 TTTTCTGTAAGATCAAGCTGGGCACTATGCCACACAGCTCCAGAGCACGTACGACCGC 783
Qy 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCCATATATTGTACATGAACAGAGTTGCTC 360
Db 784 TGCCCCATGGAGCTGGTCCGCTGCATCCGCCACATTTCTGTACAAACGAGAGCTGGTT 843
Qy 361 CGAGAACCCACATGCTAGCTCTCCAGCTGGAGCTTGTGTATGCCATGCCAGAAA 420
Db 844 CGGAGAACCCACATGCTAGCTCTCCAGCTGGAGTCTTGTGTGACGCCATGTCCAGAG 903
Qy 421 CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGCGACTGTGTGACGAGACACAGAGAA 480
Db 904 CACCTTCAGATCAACAAACGTTTGTGAGAGCTGCGCTGTATCACACAGACACGAGAAC 963
Qy 481 GAGTAAAGAGCTGTCAGCAGACTCAGAGTACTTTCATCATCTCAGTACAGAGAGCCCTG 540
Db 964 GAGCTGAAGAGCTGTCAGCAGACCCCAAGAGTACTTTCATCATCTCAGTACAGAGAGCCCTG 1023
Qy 541 AGGATCAACCTCAGTTTGGCCGCTGCCAGCTGAGCCCGAGCCCGAGGAGCGCTGAGCCGG 600
Db 1024 CGGATCAACCTCAGTTTGGCCGCTGCCAGCTGAGCCCGAGCGAGCGCATGAGCAGG 1083
Qy 601 GAGACGGCCCTCCAGCAGAGAGCTGTCTCTGGAGGCTGTGTGACGCTGAGGACAG 660
Db 1084 GAGACGGCCCTCCAGCAGAGAGCTGTCTCTGGAGGCTGTGTGACGCTGAGGACAG 1143
Qy 661 ACCTGACAGATACCGCTGGAGCTGCCAGAGACACAGAGACCCCTGAGAGCTGCTG 720
Db 1144 ACCTGACAGATACCGCTGGAGCTGCCAGAGACACAGAGAGCCCTGAGAGCTGCTG 1203
Qy 721 CGGAGCAGCAGACCATCTCTGGATCAGAGCTGATCCAGTGGAGCGGCGGAGCAG 780
Db 1204 CGGAGCAGCAGACCATCTCTGGATCAGAGCTGATCCAGTGGAGCGGAGCAGCAG 1263
Qy 781 CTGCGCGGGAACGGCGGCGCCCGAGGAGCTGGAAGCTGCTGCTGCTGCTGCTGCTG 840
Db 1264 CTGCGCGGGAACGGCGGCGCCCGAGGAGCTGGAAGCTGCTGCTGCTGCTGCTGCTG 1323
Qy 841 AAGTTGGCGGAGATCATCTGTCAGAAACGGCAGCAGATCCGAGGCTGAGACACCTCTGC 900
Db 1324 AAGTTGGCGGAGATCATCTGTCAGAAACGGCAGCAGATCCGAGGCTGAGACACCTCTGC 1383
Qy 901 CAGCAGCTGCCATCCCCCGCCAGTGGAGAGATGCTGGCGGAGGTCACCGCCACCATC 960
Db 1384 CAGCAGCTGCCATCCCCCGCCAGTGGAGAGATGCTGGCGGAGGTCACCGCCACCATC 1443
Qy 961 ACGCACATTTATCTCAGCCCTGCTGACACAGCAGCTTTCATCATTTGAGAGACGCTCCTCAG 1020
Db 1444 ACGGACATCATCTCAGCCATGCTCACAGCAGCTTTCATCATTCGAGAGACGCTCCTCAG 1503
Qy 1021 GTCTGAGAGCCACAGACCAAGTTTTCAGCCACTGTGGCCCTGTGCTGGGCGGGAAGCTG 1080

Db 1504 GTCTGAGACCCAGACCAAGTTTGCAGCCACCGTGGCGCTGCTGGTGGGGGGAAGCTG 1563
Qy 1081 AACGTGCACATGAACCCCCCCAGGTGAAGCCACCATCATCATAGTGAGCAGCAGGCCAAG 1140
Db 1564 AATGTGCACATGAACCCCCCCAGGTGAAGCCACCATCATCATAGTGAGCAGCAGGCCAAG 1623
Qy 1141 TCTCTGCTCAAGACAGGACACCCGCAATGATTACAGTGGCGAGATCTTGAACAATGTC 1200
Db 1624 TCCCTGTCTAAGAATGAGAACACCCGCAATGATTACAGCGCGGAGATCTTGAACAATGTC 1683
Qy 1201 TGGCTCATGTGAGTACCAACCAAGCCACAGGACCCCTTAGTGGCCACTTTCAGGAATATGTC 1260
Db 1684 TGGCTCATGTGAGTACCAACCAAGCCACAGGACCCCTTAGTGGCCACTTTCAGGAATATGTC 1743
Qy 1261 CTGAAACGAATTAAGAGGTGAGACCCGTCTGGGGCAGAGTGGTGACAGAGAAAATTT 1320
Db 1744 CTGAAACGAATTAAGAGGTGAGACCCGTCTGGGGCAGAGTGGTGACAGAGAAAATTT 1803
Qy 1321 ACATCTCTGTTGAATCCAGTTTCAGTGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1380
Db 1804 ACGATCTCTGTTGACTCACAGTTTCAGGCTCGGTGGAAACGAGCTGGTCTTTTCAAGTCAAG 1863
Qy 1381 ACCCTGTCTCTGCGAGTGGTGTGATCGTTTCATGGCAGCCAGGACAAATGCGACGCC 1440
Db 1864 ACCTTGTCTCTGCGAGTGGTGTGATCGTTTCATGGCAGCCAGGACAAATGCGACGCC 1923
Qy 1441 ACTGTTCTCTGGGCAATGCTTTTGCAGAGCTGGCAGGCTGCAATTTGCCGCTCGTAC 1500
Db 1924 ACTGTTCTCTGGGCAATGCTTTTGCAGAGCTGGCAGGCTGCAATTTGCCGCTCGTAC 1983
Qy 1501 AAGTGTCTGGCCACAGCTGTGTGAGCGCTCAAGATGAATTCAGGCGGAGTGCAG 1560
Db 1984 AAGTGTCTGGCCACAGCTGTGTGAGCGCTCAAGATGAATTCAGGCGTGAAGTACAG 2043
Qy 1561 AGCAACCGGGCCCTGACCAAGGAGAACTCTGTGTTCTGGCGCAGAAACTGTTCAACAAC 1620
Db 2044 AGCAACCGGGCCCTGACCAAGGAGAACTCTGTGTTCTGGCGCAGAAACTGTTCAACAAC 2103
Qy 1621 AGCAGCAGCACCCTGGAGGACTACAGTGGCTGTGTGCTCTGGTCCCAAGTCAACAG 1680
Db 2104 AGCAGCAGCACCCTGGAGGACTACAGCAGATGTCCTGCTGGTCCCAAGTCAACCGG 2163
Qy 1681 GAGATTTACAGACCGAATACACTTCTGGCAATGTTTGCAGTGTGATGGAAGTG 1740
Db 2164 GAGATTTTGCAGGACGGAATACACTTCTGGCAGTGGTTTGTGTCGCGTGAAGGATA 2223
Qy 1741 TTAATAAACAATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTTTGTAAACAAG 1800
Db 2224 TTGAAAAAANATCTCAAGCCTCATTTGGAATGATGGGGCTATCTGGGTTTCTGTAACAAG 2283
Qy 1801 CAACAGCCCCATGACTTACTGATTAACAAGCAGATGGGACCTTCTCTGAGATTCAGT 1860
Db 2284 CAACAGCCCCATGACTTACTGATTAACAAGCAGATGGGACCTTCTCTGCTGCGCTTCAGC 2343
Qy 1861 GACTCAGAAATGGCGGCATCACCATTGCTTGGAGTTTGAATCTCAGGAAGAAATGTTT 1920
Db 2344 GACTCAGAAATGGCGGCATCACCATTGCTTGGAGTTTGAATCTCAGGAAGAAATGTTT 2403
Qy 1921 TGAATCTGATGCTCTTTTACCACAGAGACTTCTTCCATCAGGTCCCTAGCGCCGCTTG 1980
Db 2404 TGGATCTGATGCTCTTTTACCACAGAGACTTCTTCTATCGGCTCCCTGCTGACCGCTG 2463
Qy 1981 GGAGACTTGAATTAACCTTATCTAGCTGTTCTGATCGGCCAAAGATGAAGTATATCTCC 2040
Db 2464 GGGGACCTGAATTAACCTTATCTAGCTGTTCTGATCGGCCAAAGATGAAGTATATCTCT 2523
Qy 2041 AATACTACACACAGTTCCTGCGAGCTGCTACTGCTAAAGCTGTTGATGATACGTG 2100
Db 2524 AAGTACTACACACCGTTCCTGAGCCCGCAACTGCGAAAGCAGCTGACGGAATACGTG 2583
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCGCGGGC 2160

Db 2584 AAGCCACAGATCAAGAGGTGGTCCCGAGTGTTCGAAATGCATCCACAGATGCTGGGAGT 2643
Qy 2161 GGCAGCGCCACGTATACATGACACAGGCCCCCTCCCCAGCTGTGTGTCCTCCAGGCTCACTAT 2220
Db 2644 G--GGCCCACTACATGATCAGGCTCCTTCCCCAGTGTGTGTCCTCCAGGCTCACTAC 2700
Qy 2221 AACATGTACCACACAGAACCCCTGACTCAGTCTCTTGACACCGATGGGACTTTCGATCTCGAG 2280
Db 2701 AACATGTACCACACACCGGAGTCCGTCCTTGATACCGATGGGACTTTCGATCTCGAA 2760
Qy 2281 GACAAATGACGTAGCGCGCGTGTGGAGAGCTCTCTGGCGCGGCAATGGACAGTCAAG 2340
Db 2761 GACACGATGACGTGGCGCGCGTGTGAGAGAGTCTTTAGCGCGGCCCATGGACAGTCAAG 2820
Qy 2341 TGGATCCCGCACGACAAATCGTGA 2364
Db 2821 TGGATCCCTCACGCACAGTCAATGA 2844

RESULT 15
AY044901
LOCUS
DEFINITION Mus musculus strain BALB/c signal transducer and activator of
transcription 5B (Stat5b) mRNA, complete cds.
ACCESSION AY044901
VERSION AY044901.1 GI:21654902
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2940)
Davoodi-Semiromi, A., Laloraya, M., Kumar, P. and She, J.-X.
Mutation detection in murine Stat gene family: identification and
characterization of a novel point mutation in DNA binding domain of
Stat5b in NOD
Unpublished
REFERENCE 2 (bases 1 to 2940)
Davoodi-Semiromi, A., Laloraya, M., Kumar, P. and She, J.-X.
Direct Submission
TITLE Submitted (10-JUL-2001) Pathology, University of Florida,
Gainesville, FL 32610, USA
JOURNAL Location/Qualifiers
FEATURES
source
1..2940
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/chromosome="11"
1..2940
/genes="Stat5b"
484..2844
/genes="Stat5b"
/notes="transcription factor"
/codon_start=1
/product="signal transducer and activator of transcription
5B"
/protein_id="AAK97791.1"
/db_xref="GI:21654903"
/translation="MAMWIAQOQLGDHALHQMOLYQHPPIEVRYHLSQWIESQAMD
SIDLNPQENIKATLLEGLVQELKKAHGVGDFLLKIKLGHVATQLQSTVDRCP
MELVRCIRHLVNEORLVREANNSSPAGSLADAMSKHLOINQTEELRLITQDTEN
ELKLOQTOEYFIHQESLRIOQAQGLQNPQERMSRETLQKQVLETLWQRE
AQTLQYRVAELAEKHOKTLQLLRQQTIIIDDELIOWKRRQQLAGGPPGSLDVQ
SWCKLAELIWNQRQITRRAHLCOOLPIPGPVEEMAEVNAITDIISALVSTFII
EKQPPVLKTQTFAATRLVGGKLVNHNPPQVATIIISQOAKSLKLNENRNDY
SGEILANNCCVMEYHQTLSAHFRNMSLKRIKSRDRGAGSVTEBKFTILFDSQSV
GNNELFQVKTLSLPVVVHGSQDNATATVLMDNAFAEPGRVPAVPDKVLWPLC
EALNMFKAQVQSRGLTKENLVFLAQKLPISSNHELDYNSVSKSOENRNLPCR
NYTFWQFQDQMEVLKKHLKHPWNGDAILGFVNKQQAHLINKPKDGTFLLRSDSEI
GTTTIAWKFDSDQSEMFNMLPFTTRDIFSIRSLADSLGDLNLYLIVFPDRPDEVYSKY
YTPVPCPEATAKAADGYVRPOIKQVVPFANASTDAGSGATYMDQAPSPVVCQAHYN
MYPNPDSVLDTDGFDEDTMDVARRVBELLGRPMDSQWIPHAQS"


```
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTCAGTTTGTGAACGATCTGCAGATGCCGGGGC 2160
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2584 AAGCCACAGATCAAGCAGGTGGTCCCGAGTTTGCAAATGCATCCACAGATGCTGGGAGT 2643
Qy 2161 GGCAGCGCCACGTACATGGACAGGCCCCCTCCCCAGCTGTGTGTCCCCAGGCTCACTAT 2220
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2644 G---GGGCCACTACATGGATCAGGCTCCTTCCCCAGTCGTGTGCCCCCTCAGGCTCACTAC 2700
Qy 2221 AACATGTACCCACAGAACCCCTGACTCAGTCCTTGACACCGATGGGGACTTCGATCTGGAG 2280
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2701 AACATGTACCCACCAACCCGACTCCGTCCTTGATACCGATGGGGACTTCGATCTGGAA 2760
Qy 2281 GACACAATGGACGTAGCGGCGGTGTGGAGGAGCTCCTGGGCGGCCCAATGGACAGTCAG 2340
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2761 GACACGATGGACGTGGGCGCGCGGTGGAGAGCTCTTAGGCGCGGCCCATGGACAGTCAG 2820
Qy 2341 TGGATCCCGCACGCACAAATCGTGA 2364
Db ||||| ||||| ||||| |||||
2821 TGGATCCCTCAGGCACAGTCATGA 2844
```

Search completed: May 17, 2005, 06:14:57
Job time : 16237 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2005, 23:15:04 ; Search time 11360 Seconds
(without alignments)
7921.121 Million cell updates/sec

Title: US-10-052-482-162
Perfect score: 2364
Sequence: 1 atggctgtgtgatacaagc.....tccgcacgcacaaatcgtga 2364

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hic:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gsal:*
- 9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1970.6	83.4	4986	3	AK045531 Mus muscu
2	956.8	40.5	1151	1	AL513644
3	923.4	39.1	1084	4	BM449533 AGENCOURT
4	862.8	36.5	1079	5	BM427445
5	841	35.6	929	7	CO648803
6	820.6	34.7	1071	4	BM450408
7	809.2	34.2	877	5	BQ932442
8	807.6	34.2	1025	5	BM401753
9	805.8	34.1	909	5	BUS08310
10	804	34.0	895	5	BQ706867
11	780	33.0	879	5	BQ706430
12	778.4	32.9	978	5	BM431739
13	767	32.4	917	5	BQ941072
14	753	31.9	991	1	AL576279
15	743.2	31.4	771	7	CF995633
16	734.6	31.1	742	5	BM4499735
17	734.2	31.1	827	7	CO740375
18	724.6	30.7	975	1	AL553220
19	717.4	30.3	755	5	BM114538
20	700.6	29.6	983	5	BM427446
21	700	29.6	886	5	BQ670478
22	664.4	28.1	756	4	BG490357
23	664	28.1	985	5	BX422302
24	658.2	27.8	730	5	BM147881

25	647.4	27.4	758	7	CF535503
26	647.4	27.4	907	1	AL521374
27	646.2	27.3	853	5	BU187856
28	642.6	27.2	1019	5	BX385196
29	638.6	27.0	715	4	BI520110
30	628.8	26.6	937	1	AL540450
31	628.4	26.6	681	4	BG394677
32	623.8	26.4	699	5	BQ939396
33	623.4	26.4	922	5	BQ939396
34	621.6	26.3	756	7	CF531926
35	621.6	26.3	1155	6	CD506834
36	617	26.1	793	2	BE545849
37	616.8	26.1	712	7	CF42767
38	614.6	26.0	706	7	CF742767
39	606.4	25.7	617	7	CF137996
40	605.8	25.6	769	4	BI415585
41	604.4	25.6	720	4	BI650503
42	604.2	25.6	828	6	CD616432
43	602	25.5	789	7	CK482172
44	589.4	24.9	811	2	BE253625
45	588.4	24.9	604	7	CN357151

ALIGNMENTS

RESULT 1	AK045531	4986 bp	mRNA	linear	HTC 03-APR-2004		
LOCUS	Mus musculus adult male corpora quadrigemina cDNA, RIKEN						
DEFINITION	full-length enriched library, clone:B230209B11 product:signal transducer and activator of transcription 5B, full insert sequence.						
ACCESSION	AK045531						
VERSION	AK045531.1	GI:26090947					
KEYWORDS	HTC; CAP trapper.						
SOURCE	Mus musculus						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
REFERENCE	1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)						
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)						
TITLE	Genome Res. 10 (10), 1617-1630 (2000)						
JOURNAL	20499374						
MEDLINE	11042159						
PUBMED	3						
REFERENCE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)						
AUTHORS	20530913						
TITLE	11076861						
JOURNAL	4						
MEDLINE	FANTOM Consortium.						
PUBMED	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)						
REFERENCE	5						
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.						
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.						
JOURNAL	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)						
MEDLINE	5						
PUBMED	The FANTOM Consortium and the RIKEN Genome Exploration Research						
REFERENCE							
AUTHORS							

Group Phase I & II Team.	
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL REFERENCE	Nature 420, 563-573 (2002)
AUTHORS	6 (bases 1 to 4986)
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp]. URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216
COMMENT	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ URL: http://location/Qualifiers
FEATURES	source 1. 4986 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:B230209B11" /db_xref="taxon:10090" /clone="B230209B11" /sex="male" /tissue_type="corpora quadrigemina" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="adult" /note="putative signal transducer and activator of transcription 5B (MGDI:103035, GB NM_011489, evidence: BLASTN, 99%, match=2574)" misc_feature 340. 2701
ORIGIN	Query Match 83.4%; Score 1970.6; DB 3; Length 4986; Best Local Similarity 90.1%; Pred. No. 0; Matches 2132; Conservative 0; Mismatches 229; Indels 4; Gaps 2; QY 1 ATGGCTGTGGATACAAAGCTCAGCAGCTCCAGAGAGAGCCCTTCATCAGATGCAAGCG 60 DB 340 ATGGCTATGTGGATACAGCTCAGCAGCTCCAGAGGCGATGCGCTTCACCATGCGAGCC 399 QY 61 TTATATGGCCAGCATTTCCCATTTAGGTGGCGCATTTATCCAGTGGATGMAAGC 120 DB 400 TTGTACGGCCAGCATTTCCCATTCGAGTGGCGACATTTATTCAGATGATCGAAGC 459 QY 121 CAAG-CATGGGACTCAGTAGATCTTTGATTAATCCACAGGAGAACATTAAGGCCACCCAGCT 179 DB 460 CAAGCCCTGGGACTCAATAGATCTTGATTAATCCACAGGAGAACATTAAGGCCACCCAGCT 519 QY 180 CCTGGAGGGGCTGGTCAGAGCTGCAGAGAGAGGACGACGACCCAGGTGGGGGAGATGG 239 DB 520 CCTGGAGGGGCTGGTCAGAGAGCTGCAGAGAGAGGCGGAGCACCAGGTGGGGGAGATGG 579 QY 240 GTTTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCACAGAACAGTATGACCG 299


```
Qy 1811 ATGACCTACTGATTAAACAGCCAGATGGGACCTTCCTCCTGAGATTCACTGACTCAGAAA 1870
Db 421 ATGACCTACTCATTAACAGCCAGATGGGACCTTCCTCCTGAGATTCACTGACTCAGAAA 480
Qy 1871 TTGGCGGCATCACCATTGCTTGGAAAGTTTGATTCTCAGGAAAGAAATGTTTTGGAAATCTGA 1930
Db 481 TTGGCGGCATCACCATTGCTTGGAAAGTTTGATTCTCAGGAAAGAAATGTTTTGGAAATCTGA 540
Qy 1931 TGCCTTTTACCACAGAGACTTCTCCATCAGGTCCTTAGCGCACCGCTTGGGAGACTTGA 1990
Db 541 TGCCTTTTACCACAGAGACTTCTCCATTCGCTTCCTTAGCGCACCGCTTGGGAGACTTGA 600
Qy 1991 ATTACCTTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATACCTCCAAATACATA 2050
Db 601 ATTACCTTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATACCTCCAAATACATA 660
Qy 2051 CACCAGTTCCTCGGAGTCTGCTACTGCTAAAGCTGTGGATAGTGAAGCCACAGA 2110
Db 661 CACCAGTTCCTCGGAGTCTGCTACTGCTAAAGCTGTGGATAGTGAAGCCACAGA 720
Qy 2111 TCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGCGGCAGCGCCA 2170
Db 721 TCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGCGGCAGCGCCA 779
Qy 2171 CGTACATGGACAGGCCCTCCCGAGCTGTGTGTCGCCAGGCTCACTATACATGTACC 2230
Db 780 CGTACATGGACAGGCCCTCCCGAGCTGTGTGTCGCCAGGCTCACTATACATGTACC 839
Qy 2231 CACAGAACCTTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAGGACACAATGG 2290
Db 840 CACAGAACCTTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAGGACACAATGG 899
Qy 2291 ACGTAGCGCGCGTGTGGAGGAGCTCTCGGCCCGCCCAATGGACAGTCAGTGGATCCCGC 2350
Db 900 ACGTAGCGCGCGTGTGGAGGAGCTCTCGGCCCGCCCAATGGACAGTCAGTGGATCCCGC 959
Qy 2351 ACGCACAAATCGTA 2364
Db 960 ACGCACAAATCGTA 973

RESULT 3
BM449533
LOCUS BM449533 1084 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT 6386129 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526703
5', mRNA sequence.
ACCESSION BM449533
VERSION BM449533.1 GI:18498573
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1084)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Cloned through the MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL12200 row: m column: 08
High quality sequence stop: 719.
Location/Qualifiers
1..1084
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db xref="taxon:9606"
/clone="IMAGE:5526703"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match 39.1%; Score 923.4; DB 4; Length 1084;
Best Local Similarity 98.0%; Pred. No. 4.4e-219;
Matches 956; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

Qy 1010 AGCCTCTCAGGTCCTGAAGACCCAGACCAAGTTTGAGCCACCTGTGCGCTGTGTGG 1069
Db 1 AGCCTCTCAGGTCCTGAAGACCCAGACCAAGTTTGAGCCACCTGTGCGCTGTGTGG 60
Qy 1070 GCGGGAAGCTGAAGCTGCACATGAACCCGCCAGGTGAAGGCCACCATCATCAGTGAGC 1129
Db 61 GCGGGAAGCTGAAGCTGCACATGAACCCGCCAGGTGAAGGCCACCATCATCAGTGAGC 120
Qy 1130 AGCAGGCCAAGTCTCTGCTCAAGAAACGAGAACACCCCAATGATTACAGTGGCGAGATCT 1189
Db 121 AGCAGGCCAAGTCTCTGCTCAAGAACGAGAACACCCCAATGATTACAGTGGCGAGATCT 180
Qy 1190 TGAACAACTGCTGCTCATGGAGTACCAACAAGCCACAGACCCCTTAGTGCCCACTTCA 1249
Db 181 TGAACAACTGCTGCTCATGGAGTACCAACAAGCCACAGACCCCTTAGTGCCCACTTCA 240
Qy 1250 GGAATATGTCCTGAAACGAATTAAGAGTTCAGACCGTCTGGGGCAGAGTCGGTGACAG 1309
Db 241 GGAATATGTCCTGAAACGAATTAAGAGTTCAGACCGTCTGGGGCAGAGTCGGTGACAG 300
Qy 1310 AAGAAAAATTTACAACTCTGTTTGAATCCAGTTCAGTGTGGTGAATAGTGGTGGTTT 1369
Db 301 AAGAAAAATTTACAACTCTGTTTGAATCCAGTTCAGTGTGGTGAATAGTGGTGGTTT 360
Qy 1370 TTCAAGTCAAGACCCCTGCTGCCAGTGTGTGTGATCGTTTCATGGCAGCAGACAACA 1429
Db 361 TTCAAGTCAAGACCCCTGCTGCCAGTGTGTGTGATCGTTTCATGGCAGCAGACAACA 420
Qy 1430 ATGGAGCGGCACCTGTTCTTGGGACAACTGTTTGGAGCCCTGGCAGGCGTGCATTTG 1489
Db 421 ATGGAGCGGCACCTGTTCTTGGGACAACTGTTTGGAGCCCTGGCAGGCGTGCATTTG 480
Qy 1490 CCGTCCCTGACAAAGTGTCTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAAATTTCAAG 1549
Db 481 CCGTCCCTGACAAAGTGTCTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAAATTTCAAG 540
Qy 1550 CCGAAGTGCAGAGCAACCGGGGCTGTGACCAAGGAAACCTCGTGTTCCTGGCGCAGAAAC 1609
Db 541 CCGAAGTGCAGAGCAACCGGGGCTGTGACCAAGGAAACCTCGTGTTCCTGGCGCAGAAAC 600
Qy 1610 TGTTCAACACAGCAGCAGCCACCTGAGGACTACAGTGGCCCTGTGTGCTCTGCTGCTCC 1669
Db 601 TGTTCAACACAGCAGCAGCCACCTGAGGACTACAGTGGCCCTGTGTGCTCTGCTGCTCC 660
Qy 1670 AGTTTCAACAGCGAGAAATTTACCAGGACGGAATTACACTTTCTGGCAATGGTTTTCAGCGTG 1729
Db 661 AGTTTCAACAGCGAGAAATTTACCAGGACGGAATTACACTTTCTGGCAATGGTTTTCAGCGTG 720
Qy 1730 TGATGGAAAGTGTAAAAAAACATCTCAAGCTCATTCGGAATGATGGGGCCA--TTTGGG 1787
Db 721 TGATGGAAAGTGTAAAAAAACATCTCAAGCTCATTCGGAATGATGGGGCCA--TTTGGG 780
Qy 1788 GTTTGTAAACAGCAACAGCCCATGACCTACTGATTAACAAGCCAGATGGGACTTCCT 1847
Db 781 GTTTGTAAACAGCAACAGCCCATGACCTACTGATTAACAAGCCAGATGGGACTTCCT 840
Qy 1848 CCTGAGATTCAAGTCACTCAGAAATTTGGCGGCATCACCATTTGCTT--GGAAGTTTGAATTC 1906
|||||
```


Tel: 2063780400
 Fax: 2063780408
 Email: cmagness@illumigen.com
 Sequenced on 2004.06.03. 715 Q20 bases. Library Preparation: Prof. Michael Katze Lab at University of Washington DNA Sequencing: Illumigen Biosciences Inc. For further information, see <http://www.macaque.org>

PCR Primers
 FORWARD: CCCTCACTAAAGGGAACAAA
 BACKWARD: CACTATAGGGCGAATGGGTA
 Insert Length: 929 Std Error: 0.00
 Plate: CL000311 row: D column: 02
 Seq primer: CCCTCACTAAAGGGAACAAA
 POLYA=No.

FEATURES

source

Location/Qualifiers

1. 929

/organism="Macaca mulatta"
 /mol_type="mRNA"
 /strain="Indian"
 /db_xref="taxon:9544"
 /clone="IBI0W:22257"
 /sex="male"
 /tissue_type="blood"
 /cell_type="PBMC"
 /dev_stage="adult"
 /lab_host="Electromax DH10B"
 /clone_lib="Katze MMPB2"
 /note="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I; Created from CloneMiner cDNA Library Construction kit (catalog #18249-029)"

ORIGIN

Query Match 35.6%; Score 841; DB 7; Length 929;
 Best Local Similarity 96.3%; Pred. No. 1.6e-198;
 Matches 893; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 1268 GAATTAGAGGTGACCGTGGGGCAGAGTCGGTGACAGAGAAATTTACAATCC 1327
 DB 3 GGATTAGAGATCAGACCGTGGGGCAGAGTCAGTGACAGAGAAATTTACAATCC 62

QY 1328 TGTGTGAATCCAGTTCAGTGTGGTGAATGAGTGGTTTTCAAGTCAAGACCCGTG 1387
 DB 63 TGTGTGAATCCAGTTCAGTGTGGTGAATGAGTGGTTTTCAAGTCAAGACCCGTG 122

QY 1388 CCCTGCCAGTGGTGTGATGCTTATGCGACGCCAGGACAAACAATGCGAGCCCACTGTTC 1447
 DB 123 CCCTGCCAGTGGTGTGATGCTTATGCGACGCCAGGACAAACAATGCGAGCCCACTGTTC 182

QY 1448 TCTGGGCAATGCTTTTGCAGAGCTGCGAGGTTGCCATTTGCCGTGCCTGACAAAGTGC 1507
 DB 183 TCTGGGCAATGCTTTTGCAGAGCTGCGAGGTTGCCATTTGCCGTGCCTGACAAAGTGC 242

QY 1508 TGTGGCCAAGTGTGTGAGCGCTCAACATGAAATTAAGGCCGAAAGTGACAGCAACC 1567
 DB 243 TGTGGCCAAGTGTGTGAGCGCTCAACATGAAATTAAGGCCGAAAGTGACAGCAACC 302

QY 1568 GGGGCTGTACCAAGAGAACCTCGTGTTCCTGGGCGCAGAAACTGTTCAACAGCAGCA 1627
 DB 303 GGGGCTGTACCAAGAGAACCTCGTGTTCCTGGGCGCAGAAACTGTTCAACAGCAGCA 362

QY 1628 GCCACCTGGAGACTACAGTGGCTGTCTGTGTCTGTCTCCAGTTCAACAGGAGAAAT 1687
 DB 363 GCCACCTGGAGACTACAGTGGCTGTCTGTGTCTGTCTCCAGTTCAACAGGAGAAAT 422

QY 1688 TACCAGGACGGAATTAACATTTCTGGCAATGGTTGACGGTGTGATGGAAGTGTAAAAA 1747
 DB 423 TACCAGGACGGAATTAACATTTCTGGCAGTGGTTGACGGTGTGATGGAAGTGTAAAAA 482

QY 1748 AACATCTCAAGCCTCATTTGGAATGATGGGCCATTTTGGGGTTTGTAAACAAGCAAGG 1807
 DB 483 AACATCTCAAGCCTCATTTGGAATGATGGGCCATTTTGGGGTTTGTAAACAAGCAAGG 542

QY 1808 CCCATGACCTACTGATTAAACAAGCCAGATGGGACCTTCTCCTCAGATTTCAGTCACTCAG 1867

DB 543 CCCATGACCTACTCATTAACAGCCAGATGGGACCTTCTCCTCAGATTTCAGTCACTCAG 602

QY 1868 AAATTGGCGGATCACCATTTGCTTGGAAAGTTTGAATCTCAGGAAAGAAATTTGGGAATC 1927

DB 603 AAATTGGCGGATCACCATTTGCTTGGAAAGTTTGAATCTCAGGAAAGAAATTTGGGAATC 662

QY 1928 TGATGCCCTTTTACCCACAGAGACTTCTCCATCAGTCCCTAGCCGACCGCTTGGGAGACT 1987

DB 663 TGATGCCCTTTTACCCACAGAGACTTCTCCATTCGGTCCCTAGCCGACCGCTTGGGAGACT 722

QY 1988 TGAATTACCTTATCTACGTGTTCTCGATCGGGCAAAAGATGAAGTATATCTCCAAAATACT 2047

DB 723 TGAATTACCTTATCTACGTGTTCTCGATCGGGCAAAAGATGAAGTATATCTCCAAAATACT 782

QY 2048 ACACACCAGTTCCCTCGAGTCTGCTACTGCTAAAG-CTGTGTGATGATACGTGAAGCCA 2106

DB 783 ACACACCAGTTCCCTCGAGTCTGCTACTGCTAAAGCCCGTTGATGATACGTGAAGCCA 842

QY 2107 CAGATCAAGCAAGTGG-TCCCTGAGTTTGTGAACGCACTGCAGATGCCGG--GGCGGC 2163

DB 843 CAAATCAAGCAAGTGGTTCCCTGAAATTTGTGAACCCCTTCTGCAAAATGCGGGGAGCGGGC 902

QY 2164 AGCGCCACGTACATGGACCAAGGCCCC 2190

DB 903 AGCGCCCTTACCTGGACCAAGGCCCC 929

RESULT 6
 BM450408 1071 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT_6394143 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5494762
 DEFINITION S', mRNA sequence.
 ACCESSION BM450408
 VERSION BM450408.1 GI:18499448
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1071)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM12120 row: j column: 11
 High quality sequence stop: 633.
 Location/Qualifiers
 1. 1071
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5494762"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_67"

FEATURES
 source
 1. 1071
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5494762"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_67"

ORIGIN
 Query Match 34.7%; Score 820.6; DB 4; Length 1071;
 Best Local Similarity 96.2%; Pred. No. 2e-193;
 Matches 862; Conservative 0; Mismatches 30; Indels 4; Gaps 2;

```
QY 1284 CCGTCTGGGGCAGAGTCGGTGACAGAGAAAAATTTACATCTCTGTTGTAATCCAGTT 1343
Db 1 CCGTCTGGGGCAGAGTCGGTGACAGAGAAAAATTTACATCTCTGTTGTAATCCAGTT 60
QY 1344 CAGTGTGGTGAATAGAGCTGGTTTTTCAAGTCAAGACCTGTGCTCCGCGAGTGGTGGT 1403
Db 61 CAGTGTGGTGAATAGAGCTGGTTTTTCAAGTCAAGACCTGTGCTCCGCGAGTGGTGGT 120
QY 1404 GATCGTTTCATGCGCAGGAGCAACAATGCGAGCGCACTGTTCTCTGGGCAATGCTTT 1463
Db 121 GATCGTTTCATGCGCAGGAGCAACAATGCGAGCGCACTGTTCTCTGGGCAATGCTTT 180
QY 1464 TGCAGAGCTGCGCAGGCTGCAATTTGCGTGTGCTGACAAAGTGTGCTGGCCACAGCTGTG 1523
Db 181 TGCAGAGCTGCGCAGGCTGCAATTTGCGTGTGCTGACAAAGTGTGCTGGCCACAGCTGTG 240
QY 1524 TGAGCGCTCAACATGAATTAAGGCGGAGTGCAGAGCAACCGGGGCTGTGACCAAGGA 1583
Db 241 TGAGCGCTCAACATGAATTAAGGCGGAGTGCAGAGCAACCGGGGCTGTGACCAAGGA 300
QY 1584 GAACCTGTGTTCTGGCGCAGAACTGTTCAACAGCAGCAGCCACCTGGAGGACTA 1643
Db 301 GAACCTGTGTTCTGGCGCAGAACTGTTCAACAGCAGCAGCCACCTGGAGGACTA 360
QY 1644 CAGTGGCTGTCTGTGCTGCTGCTCCAGTTTCAACAGGAGAAATTTACAGGACGGAATTA 1703
Db 361 CAGTGGCTGTCTGTGCTGCTGCTCCAGTTTCAACAGGAGAAATTTACAGGACGGAATTA 420
QY 1704 CACTTTCTGGCAATGGTTTGACGGTGTGATGAAGTGTGTTAAAGAAATCTCAAGCCTCA 1763
Db 421 CACTTTCTGGCAATGGTTTGACGGTGTGATGAAGTGTGTTAAAGAAATCTCAAGCCTCA 480
QY 1764 TTGGAATGATGGGCGCAATTTGGGGTGTGTTAAACAGCAACGCGCCATGACCTACTGAT 1823
Db 481 TTGGAATGATGGGCGCAATTTGGGGTGTGTTAAACAGCAACGCGCCATGACCTACTGAT 540
QY 1824 TAACAGCCAGATGGGACCTTCTCTGAGTTCAGTGAATTCAGAGAAATTTGGGGATCAC 1883
Db 541 TAACAGCCAGATGGGACCTTCTCTGAGTTCAGTGAATTCAGAGAAATTTGGGGATCAC 600
QY 1884 CATTCCTTGGAGTTGATTTCTCAGAGAAAGATGTTTGGATCTGATGCTTTTACCAC 1943
Db 601 CATTCCTTGGAGTTGATTTCTCAGAGAAAGATGTTTGGATCTGATGCTTTTACCAC 660
QY 1944 CAGAGACTTCTCAGTCCCTAGCCGACCGCTTGGAGACTTGAATTAATCTTATCTA 2003
Db 661 CAGAGACTTCTCAGTCCCTAGCCGACCGCTTGGAGACTTGAATTAATCTTATCTA 720
QY 2004 CGTGTTCCTGATCGGCCAAGATGAAGTATCTCCTCAATTAATACACACAG-TTCCTT 2062
Db 721 CGTGTTCCTGATCGGCCAAGATGAAGTATCTCCTCAATTAATACACACAGTTCCTT 780
QY 2063 GCGAGTCTGCTACTGCTAAGCTGTGATGATGATGATGATGATGATGATGATGATGATG 2122
Db 781 GCGAGTCTGCTACTGCTAAGCTGTGATGATGATGATGATGATGATGATGATGATGATG 840
QY 2123 TCCC---TGAGTTTGTGAACGATCTGACAGATGCGCGGGGGCGGACGCCACGTCAC 2175
Db 841 GCCCTCGAAATTTGTTGACCGCATCTGCCAAATGTCGCGGGGCGGCAACCCCCAC 896
```

```
RESULT 7
BQ932442
LOCUS BQ932442 8854276 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6473729
DEFINITION BQ932442 877 bp mRNA linear EST 21-AUG-2002
5', mRNA Sequence.
ACCESSION BQ932442
VERSION BQ932442.1 GI:22347796
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 877)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14009 row: 1 column: 18
High quality sequence stop: 662.
FEATURES
location/Qualifiers
1..877
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6473729"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
ORIGIN
Query Match 34.2%; Score 809.2; DB 5; Length 877;
Best Local Similarity 98.9%; Pred. No. 1.4e-190;
Matches 825; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 361 CGAAGAACCAATAGTGTAGCTCTCCAGCTGGAAGCCTTGTGATGCATGTCCAGAAA 420
Db 1 CGAAGAACCAATAGTGTAGCTCTCCAGCTGGAAGCCTTGTGATGCATGTCCAGAAA 60
QY 421 CACCTCCAGATCAACAGACGCTTTGAGGAGCTGGACTGGTCCAGCAGACACAGAGAT 480
Db 61 CACCTCCAGATCAACAGACGCTTTGAGGAGCTGGACTGGTCCAGCAGACACAGAGAT 120
QY 481 GAGTAAAAAGCTGACAGACTCAGAGTACTTTCATCCAGTACAGAGAGCCTG 540
Db 121 GAGTAAAAAGCTGACAGACTCAGAGTACTTTCATCCAGTACAGAGAGCCTG 180
QY 541 AGGATCCAGCTCAGTTTGGCCCGCTGCCAGCTGAGCCCCCAGAGAGCTGTGAGCCGG 600
Db 181 AGGATCCAGCTCAGTTTGGCCCGCTGCCAGCTGAGCCCCCAGAGAGCTGTGAGCCGG 240
QY 601 GAGAGCGCCCTCCAGCAGAACAGGTGTCTCTGGAGGCTGTTGACGGTGGAGCAG 660
Db 241 GAGAGCGCCCTCCAGCAGAACAGGTGTCTCTGGAGGCTGTTGACGGTGGAGCAG 300
QY 661 ACCTGACAGCTACCGCTGGAGCTGCCAGAGACACAGAACCTCTGAGCTGCTG 720
Db 301 ACCTGACAGCTACCGCTGGAGCTGCCAGAGACACAGAACCTCTGAGCTGCTG 360
QY 721 CGAAGCAGCAGACCATCATCTGATGACAGCTGATCCAGTGAAGCGGCGGAGCAG 780
Db 361 CGAAGCAGCAGACCATCATCTGATGACAGCTGATCCAGTGAAGCGGCGGAGCAG 420
QY 781 CTGGCGGGAAACGGGGGGCCCCCGAGGAGCCTTGGAGCTGTACGTCTCTGGTGTGAG 840
Db 421 CTGGCGGGAAACGGGGGGCCCCCGAGGAGCCTTGGAGCTGTACGTCTCTGGTGTGAG 480
QY 841 AAGTTGGCGGAGATCATCTGCGAGAACCGGCGAGCAGATCCGCGAGGCTGAGCACCTCTGCG 900
Db 481 AAGTTGGCGGAGATCATCTGCGAGAACCGGCGAGCAGATCCGCGAGGCTGAGCACCTCTGCG 540
QY 901 CAGCAGCTGCCCATCCCCCGGCCAGAGTGTGGCGAGGATGTGGCGAGGTCAACGCCACCATC 960
Db 541 CAGCAGCTGCCCATCCCCCGGCCAGAGTGTGGCGAGGATGTGGCGAGGTCAACGCCACCATC 600
```

```

Qy 961 ACGGACATTATCTCAGCCCTGGTGACACAGACGTTTCATTCATTGAGACGACCTCCCTCAG 1020
Db 601 ACGGACATTATCTCAGCCCTGGTGACACAGACGTTTCATTCATTGAGACGACCTCCCTCAG 660
Qy 1021 GTCCTGAAGACCCAGACACCAAGTTTGACGCCACTGTGCGCTGCTGGTGGGCGGAAGCTG 1080
Db 661 GTCCTGAAGACCCAGACACCAAGTTTGACGCCACTGTGCGCTGCTGGTGGGCGGAAGCTG 720
Qy 1081 AACGTGCATGAACACCCCCCGAGGTGAAGGCCACCATCATCAGTGAAGCAGACGCCCAAG 1140
Db 721 AACGTGCATGAACACCCCCCGAGGTGAAGGCCACCATCATCAGTGAAGCAGACGCCCAAG 780
Qy 1141 TCTCTGCTCAAGAACGAGAACCCCGCAATCATTTACAGT-GCGGAGATCTTGAA 1193
Db 781 TCTCTGCTCAAGAACGAGAACCCCGCAATCATTTACAGTGGGCGAGATCTTGAA 834

RESULT 8
BX401753
LOCUS BX401753 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-APR-2004
DEFINITION Homo sapiens cDNA clone CS0DL008YC13 5-PRIME, mRNA sequence.
ACCESSION BX401753
VERSION BX401753.2 GI:46835520
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1025)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization.
JOURNAL Unpublished (2001)
COMMENT On May 13, 2003 this sequence version replaced gi:30610754.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
8244.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DL008AB07QPI&c=8244.r.
FEATURES
Location/Qualifiers
1..1025
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL008YC13"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dr)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 34.2%; Score 807.6; DB 5; Length 1025;
Best Local Similarity 97.2%; Pred. No. 3.5e-190;
Matches 846; Conservative 6; Mismatches 15; Indels 3; Gaps 3;
Qy 1 ATGGCTGTGGATCAAGCTCAGAGCTCCAGGAGAGCCCTTCATCAGATGCAAGCG 60
Db 148 ATGGCTGTGGATCAAGCTCAGAGCTCCAGGAGAGCCCTTCATCAGATGCAAGCG 207
Qy 61 TTATATGCCAGCATTTTCCCATTTAGGTGGCGCATTTATTTATCCAGTGGATTGAAAGC 120

```

```

Db 208 TTATATGCCAGCATTTTCCCATTTAGGTGGCGCATTTATTTATCCAGTGGATTGAAAGC 267
Qy 121 CAAGCATGGGACTCAGTAGATCTTTGATAATCCACAGGAGAACATTAAAGGCCACCCAGCTC 180
Db 268 CAAGCATGGGACTCAGTAGATCTTTGATAATCCACAGGAGAACATTAAAGGCCACCCAGCTC 327
Qy 181 CTGAGAGGCGCTGGTGGCAGGAGCTGCAGAAGAAGGCAGACACAGGTGGGGAAGATGGG 240
Db 328 CTGAGAGGCGCTGGTGGCAGGAGCTGCAGAAGAAGGCAGACACAGGTGGGGAAGATGGG 387
Qy 241 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACAGTATGACCGC 300
Db 388 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACAGTATGACCGC 447
Qy 301 TGCCCCATGAGCTGGTCCGCTGCATCCGCCATATATTGTACATGAACAGAGGTTGGTC 360
Db 448 TGCCCCATGAGCTGGTCCGCTGCATCCGCCATATATTGTACATGAACAGAGGTTGGTC 507
Qy 361 CGAGAAGCCAAATAGTGTAGCTCTCCAGCTGGGAAGCCTTGCTGATGCATGTCCAGAAA 420
Db 508 CGAGAAGCCAAATAGTGTAGCTCTCCAGCTGGGAAGCCTTGCTGATGCATGTCCAGAAA 567
Qy 421 CACCTCCAGATCAACACAGAGCTTTTGGAGAGCTGCGACTGTGTCACGACGAGCACAGAGAAT 480
Db 568 CACCTCCAGATCAACACAGAGCTTTTGGAGAGCTGCGACTGTGTCACGACGAGCACAGAGAAT 627
Qy 481 GAGTTAAAAAGCTGCAGCAGACTCAGAGTAGTACTTTCATCATCTCCAGTACCAGGAGAGCTG 540
Db 628 GAGTTAAAAAGCTGCAGCAGACTCAGAGTAGTACTTTCATCATCTCCAGTACCAGGAGAGCTG 687
Qy 541 AGGATCCAAGCTCAGTTTGGCCCGCTGCCAGCTGAGCCCCAGGAGCGCTCTCAGCGCG 600
Db 688 AGGATCCAAGCTCAGTTTGGCCCGCTGCCAGCTGAGCCCCAGGAGCGCTCTCAGCGCG 747
Qy 601 GAGACGCGCTCCACAGAGAAGCAGGTGCTCTGTGAGGCGCTGGTTGCGAGCGGTGAGGCAAG 660
Db 748 GAGACGCGCTCCACAGAGAAGCAGGTGCTCTGTGAGGCGCTGGTTGCGAGCGGTGAGGCAAG 807
Qy 661 ACATCGACGAGTACCCGCTGGAGCTCCCGAGAGAACACACAGAGACCTTCAGCTGCTG 720
Db 808 ACATCGACGAGTACCCGCTGGAGCTGGCGAGAAGCACACAGAGAACCTTCAGCTGCTG 867
Qy 721 CGGAAGCAGACAGACCATCATCCT-GGATGACGAGCTGATCCAGTGAAGCGCGCGCAGCA 779
Db 868 CGGAAGCMRGNACCAATCATCTCKGATGACGAGCTGATCCAGTGGAGCGCGCGCA-CA 926
Qy 780 GCTGCGCGGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
Db 927 GCTGCGCGGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 986
Qy 840 GAGTTGGCGGAGATCATCTGGCAGAACCG 869
Db 987 G-AGTTGGCGGAGATCATCTTGGMRARCCGG 1015

RESULT 9
BX508310
LOCUS BX508310 Homo sapiens 909 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT 10094472 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6502375
5', mRNA sequence.
ACCESSION BU508310
VERSION BU508310.1 GI:22814543
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 909)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-romail.nih.gov

```

Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14058 row: f column: 08
High quality sequence stop: 638.

FEATURES
source
1. .909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6502375"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb."

ORIGIN
Query Match 34.1%; Score 805.8; DB 5; Length 909;
Best Local Similarity 97.9%; Pred. No. 9.7e-190;
Matches 890; Conservative 0; Mismatches 12; Indels 7; Gaps 7;

QY 1048 GCCACTGTCGCTGCTGCTGGGGGAAAGCTGAACGTGCACATGAACCCCCCGAGTG 1107
DB 1 GCCACTGTCGCTGCTGCTGGGGGAAAGCTGAACGTGCACATGAACCCCCCGAGTG 60

QY 1108 AAGGCCACCATCATCAGTGAGCAGCAGCCCAAGTCTCTGCTCAAGAACGAGAACACCCGC 1167
DB 61 AAGGCCACCATCATCAGTGAGCAGCAGCCCAAGTCTCTGCTCAAGAACGAGAACACCCGC 120

QY 1168 AATGATTACAGTGGCGAGATCTGAACAACTGCTGCTCATGGAGTACCCCAAGCCACA 1227
DB 121 AATGATTACAGTGGCGAGATCTGAACAACTGCTGCTCATGGAGTACCCCAAGCCACA 180

QY 1228 GGCACCTTAGTGCCCACTTACGGAATATGCTCTGAAACGAATTAAGAGTGCAGACCGT 1287
DB 181 GGCACCTTAGTGCCCACTTACGGAATATGCTCTGAAACGAATTAAGAGTGCAGACCGT 240

QY 1288 CGTGGGCGAGTGGTGCACAGAGAAATTTACAATCCTGTTTGAATCCAGTTCACT 1347
DB 241 CGTGGGCGAGTGGTGCACAGAGAAATTTACAATCCTGTTTGAATCCAGTTCACT 300

QY 1348 GTTGTGGAATGAGCTGGTCTTCAAGTCAAGACCTGCTCCCTGCCAGTGGTGTGATC 1407
DB 301 GTTGTGGAATGAGCTGGTCTTCAAGTCAAGACCTGCTCCCTGCCAGTGGTGTGATC 360

QY 1408 GTTCATGCAGCAGGACCAATGCGACGGCCACTGTTCTCTGGGACAATGCTTTTGCA 1467
DB 361 GTTCATGCAGCAGGACCAATGCGACGGCCACTGTTCTCTGGGACAATGCTTTTGCA 420

QY 1468 GAGCTGCGAGGTCCTATTTGCTGCTGCAAAAGTCTGTGCGCACAGCTGTGAG 1527
DB 421 GAGCTGCGAGGTCCTATTTGCTGCTGCAAAAGTCTGTGCGCACAGCTGTGAG 480

QY 1528 GGGCTCAACATGAAATTCAGGCGCAAGTGCAGAGCAACCGGGGCTGCACCAAGGAGAAC 1587
DB 481 GGGCTCAACATGAAATTCAGGCGCAAGTGCAGAGCAACCGGGGCTGCACCAAGGAGAAC 540

QY 1588 CTCGTGTTCTGGCGCAGAACTGTTCAACACAGCAGCAGCCACTGGAGACTACAGT 1647
DB 541 CTCGTGTTCTGGCGCAGAACTGTTCAACACAGCAGCAGCCACTGGAGACTACAGT 600

QY 1648 GGCCTGCTGTGCTCTGTGCTCCAGTTCAACAGGAGAAATTTACAGGACGGAATTTACACT 1707
DB 601 GGCCTGCTGTGCTCTGTGCTCCAGTTCAACAGGAGAAATTTACAGGACGGAATTTACACT 660

QY 1708 TTCTGGCAATGTTTGGCGTGTGATGGAAGTGTAAAAAATCTCAAGCTTCATGG 1767
|||||

Db 661 TTCTGGCAATGTTTGGCGTGTGATGGAAGTGTAAAAAATCTCAAGCTTCATGG 719

QY 1768 AATGATGGGGCCA-TTTTGGGGTTTGTAAACAAGCAACAGG-CCCATGACCTACTGATTA 1825
|||||

Db 720 AATGATGGGGCCATTTTGGGTTTGTAAACAAGCAACAGGCCCCCATCTACTCATTA 779
|||||

QY 1826 ACAAGCCAGATGGGACCTTCTCT-CCTGAGATTAGTCACTCAG-AAATTGGCGGCATCAC 1883
|||||

Db 780 ACAAGCCAGATGGGACCTTCTCTCCTGAGATTCACTGACTCAGAAAAATTTGGCGCATCAC 839
|||||

QY 1884 CA-TTGCTTGGAGTTTGAATCTCAGGAAGAAATTTTGGAAAT-CTGATCCCTTTTACC 1941
|||||

Db 840 CATTTGCTTGGAGTTTGAATCTTTCGAAAAAGAAAGTTTGGAAATCCGAGGCCCTTTTACC 899
|||||

QY 1942 ACCAGAGAC 1950
|||||

Db 900 ACCCAAAAC 908
|||||

RESULT 10
LOCUS B0706867
DEFINITION AGENCOURT_8353060 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6277975
5', mRNA sequence.
ACCESSION B0706867
VERSION B0706867.1 GI:21845766
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 895)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2465 row: d column: 08
High quality sequence stop: 712.

FEATURES
source
1. .895
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6277975"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 113"
/note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 34.0%; Score 804; DB 5; Length 895;
Best Local Similarity 99.4%; Pred. No. 2.7e-189;
Matches 807; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1553 AAGTGCAGACAACCGGGCCTGACCAAGGAGAACCTGTTCTCTGGCGCAGAACTGT 1612
DB 1 AAGTGCAGACAACCGGGCCTGACCAAGGAGAACCTGTTCTCTGGCGCAGAACTGT 60

QY 1613 TCACAACAGCAGCAGCAGCCACCTGGAGGACTACAGTGGCCTGTCTGTCTGCTGCCAGT 1672
|||||

Db 661 CTCACATAACATGTATACACAGAACCTCTAGTCTTGACACCGATGGGACTTCG 720
Qy 2273 ATCTGGAGACACATAGACCTAGCGCGGTGTGGAGAGCTCTGGCGGCCCAATGG 2332
Db 721 ATCTGGAGGACAC-ATGGACGTAAACGCGCGGTGTGGAGAGCTCTGGCGGCCCAATGG 779
Qy 2333 ACAGTCA-GTGGATCCCGCAGCACAATCGTGA 2364
Db 780 ACAGTCAGTGGATCCCGAACGCACAATCGTGA 812
RESULT 12
LOCUS BX431739 978 bp mRNA linear EST 04-MAY-2004
DEFINITION BX431739 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
CSODM003YD01 5-PRIME, mRNA sequence.
ACCESSION BX431739
VERSION BX431739.1 GI:30783021
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 978)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE
Email: secre@genoscope.cns.fr, Web: www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 8244.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?B=CSOBAG063ZA12_CS06041_1&c=8244.r

FEATURES
source

Location/Qualifiers
1..978
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODM003YD01"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match 32.9%; Score 778.4; DB 5; Length 978;
Best Local Similarity 89.6%; Pred. No. 6.8e-183;
Matches 881; Conservative 0; Mismatches 94; Indels 8; Gaps 4;
Qy 590 GTCTGAGCGGAGAGCGGCCCTCCAGCAGACAGCTGTCTGGAGCGCTGTTGCAGC 649
Db 1 GTCTGAGCGGAGAGCGGCCCTCCAGCAGACAGCTGTCTGGAGCGCTGTTGCAGC 60
Qy 650 GTGAGCAGACAGACTGCGAGCAGTACCGGTGGAGCTGCCGAGAGCAGCAGAGACCC 709
Db 61 GTGAGCAGACAGACTGCGAGCAGTACCGGTGGAGCTGCCGAGAGCAGCAGAGACCC 120
Qy 710 TGCAGCTGTCTGGAGAGCAGCAGACCATCATCTGTGATGACGAGCTGATCCAGTGGAGC 769
Db 121 TGCAGCTGTCTGGAGAGCAGCAGACCATCATCTGTGATGACGAGCTGATCCAGTGGAGC 180

Qy 770 GGCGGACAGCACTGGCCGGGAAACGGCGGGCCCCCGAGGGCAGCTGGAACGCTGCTACAGT 829
Db 181 GGCGGACAGCACTGGCCGGGAAACGGCGGGCCCCCGAGGGCAGCTGGAACGCTGCTACAGT 240
Qy 830 CCTGGTGTGAGAAAGTTGGCGGAGATCATCTGGCAGAACCGGCAGCAGATCCGACGGCTG 889
Db 241 CCTGGTGTGAGAAAGTTGGCGGAGATCATCTGGCAGAACCGGCAGCAGATCCGACGGCTG 300
Qy 890 AGCACCTCTGCGCAGCAGCTGCCCATCCCGGCCCCAGTGGAGGAGATGCTGGCCGAGGTCA 949
Db 301 AGCACCTCTGCGCAGCAGCTGCCCATCCCGGCCCCAGTGGAGGAGATGCTGGCCGAGGTCA 360
Qy 950 AGCGCCACCATCACCGACATTATCTCAGCCCTGGTGACCAAGCGCTTCATCATTTGAGAAC 1009
Db 361 AGCGCCACCATCACCGACATTATCTCAGCCCTGGTGACCAAGCGCTTCATCATTTGAGAAC 420
Qy 1010 AGCCTCTCAGGTCTGTGAGAGCCAGACCAAGTTTGCAGCCACTGTGCGCTGCTGGTGG 1069
Db 421 AGCCTCTCAGGTCTGTGAGAGCCAGACCAAGTTTGCAGCCACTGTGCGCTGCTGGTGG 480
Qy 1070 GCGGGAAGCTGAACGCTGCACATGAACCCCCCGGAGTGAAGGCCACCATCATCAGTGAAGC 1129
Db 481 GCGGGAAGCTGAACGCTGCACATGAATCCCCCGGAGTGAAGGCCACCATCATCAGTGAAGC 540
Qy 1130 AGCAGGCCCAAGTCTCTGCTCAAGAACGAGAACACCCGCAATGATTACAGTGGCGAGATCT 1189
Db 541 AGCAGGCCCAAGTCTCTGCTTAAATAATGAGAACACCCGCAACGAGTGCAGTGGTGAATCC 600
Qy 1190 TGAACAACTGCTGGTCTATGAGATACCAAGCCACAGGCCACCCCTTAGTGCCTTCA 1249
Db 601 TGAACAACTGCTGGTCTATGAGATACCAAGCCACAGGCCACCCCTTAGTGCCTTCA 660
Qy 1250 GGAATATGCTCCCTGAAACGAATTAAGAGGTGAGACCGCTGCGGGCAGAGTCCGTGACAG 1309
Db 661 GGAATATGCTCCCTGAAAGAGGTGAGACCGCTGCGGGCAGAGTCCGTGACAG 720
Qy 1310 AAGAAAAATTTACAAATCTCTTTGAATCCCAAGTTCAGTGTGGTGGAAATGAGCTGGTTT 1369
Db 721 AGGAGAAGTTCACAGTCCCTGTTGAGTCTCAGTTCAGTGTGGCAGCAATGAGCTGTGT 780
Qy 1370 TTCAAGTCAAGACCCCTGCTCCCTGCGCAGGTGGTGTGATCGTTTCATGCGCAGGCAACA 1429
Db 781 TTCAGGTGAAGACTCTGTNCCCTACCTGTGGTGTGTCATCGTCACGCGGAGCCAGGCCAC 840
Qy 1430 ATGCGAGCGCCACATGTTCTCTGGGCAATGCTTTTGGCAGAGCTCGCAGGGGTGCCATTGG 1489
Db 841 ATGCCACCGGCTACTGTGCTGTGGGAC-ATGCCCTTGTCTGAGCC-GGCAGGGTGCCATTGG 898
Qy 1490 CCGTGCCTTGACAAAGTGTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAAATTCAGG 1549
Db 899 CCGTGCCTTGACAAA---TGCTGGGNCGCCACTGTGTGAGCGCTCACATGAAATCAAGGC 954
Qy 1550 CCGAAGTGCAGAGCAACCGGGGC 1572
Db 955 CC--AATGCAAAAGAAACCGGGGC 975
RESULT 13
LOCUS BQ941072
DEFINITION BQ941072 8853726 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6374059
5', mRNA sequence.
ACCESSION BQ941072
VERSION BQ941072.1 GI:22356550
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 917)
AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: csapbs-remail.nih.gov
Tissue Procurement: DCID/DBP/Gazdar
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution by: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2552 row: g column: 20
High quality sequence stop: 591.
Location/Qualifiers

FEATURES
source 1. 917
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6374059"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_18"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 32.4%; Score 767; DB 5; Length 917;
Best Local Similarity 95.5%; Pred. No. 4.7e-180; Indels 5; Gaps 5;
Matches 843; Conservative 0; Mismatches 35;

Qy 448 GAGCTGGACCTGGTCAACGAGACACAGAGATGAGTTAAAGCTGCAGACTCAG 507
Db 1 GAGCTGGACCTGGTCAACGAGACACAGAGATGAGTTAAAGCTGCAGACTCAG 60

Qy 508 GAGTACTTCATCTCAGTACAGAGAGCTGAGGATCCAGCTGATTTGGCCCGCTG 567
Db 61 GAGTACTTCATCTCAGTACAGAGAGCTGAGGATCCAGCTGATTTGGCCCGCTG 120

Qy 568 GCCAGCTGAGCCCGCCAGAGCGTCTAGCGGGAGAGCGCCCTCCAGCAGAACAGCTG 627
Db 121 GCCAGCTGAGCCCGCCAGAGCGTCTAGCGGGAGAGCGCCCTCCAGCAGAACAGCTG 180

Qy 628 TCTCTGGAGGCTGGTGTGACGCTGAGCAGACACTGACGAGTACCGGCTGGAGCTG 687
Db 181 TCTCTGGAGGCTGGTGTGACGCTGAGCAGACACTGACGAGTACCGGCTGGAGCTG 240

Qy 688 CCCGAGAGCACCAGAGACCTGAGCTGTGCGGAGAGCAGACCATCATCTGTGAT 747
Db 241 CCCGAGAGCACCAGAGACCTGAGCTGTGCGGAGAGCAGACCATCATCTGTGAT 300

Qy 748 GACGAGCTGATCCAGTGAAGCGCGCAGAGCTGGCGGGAACCGCGGCCCGCCGAG 807
Db 301 GACGAGCTGATCCAGTGAAGCGCGCAGAGCTGGCGGGAACCGCGGCCCGCCGAG 360

Qy 808 GGCAGCTGGACGTGCTACAGTCTCTGTGTGAGAAAGTTGGCGGAGATCATCTGGCAGAAC 867
Db 361 GGCAGCTGGACGTGCTACAGTCTCTGTGTGAGAAAGTTGGCGGAGATCATCTGGCAGAAC 420

Qy 868 CGGAGCAGATCCGAGGGCTGAGACCTCTGCGCAGAGCTGCCCATCCCCGGGCCAGTG 927
Db 421 CGGAGCAGATCCGAGGGCTGAGACCTCTGCGCAGAGCTGCCCATCCCCGGGCCAGTG 480

Qy 928 GAGGAGATGCTGGCGGAGGTCAAGCCACCATCAGGACATATCTCAGCCCTGTGACC 987
Db 481 GAGGAGATGCTGGCGGAGGTCAAGCCACCATCAGGACATATCTCAGCCCTGTGACC 540

Qy 988 AGCAGCTTCATCTTGAAGACAGCCTCTCCTCAGGTCTCTGAAGACCCAGCAAGTTTGCA 1047
Db 541 AGCAGCTTCATCTTGAAGACAGCCTCTCCTCAGGTCTCTGAAGACCCAGCAAGTTTGCA 600

Qy 1048 GCCACTGTGCGCTGTGTGGGCGGCGGAGCTGAACGTGCACATGAACCCCCCAGGT 1106
Db 601 GCCACTGTGCGCTGTGTGGGCGGAGCTGAACGTGCACATGAACCCCCCAGGT 660

Qy 1107 GAAGGCCACCATCATCTAGTGAGCAGAGCGCAAGTCTCTGCTCAAGAACGAGAAC-CCC 1165
Db 661 GAAGGCCACCATCATCTAGTGAGCAGAGCGCAAGTCTCTGCTCAAGAACGAGAAC-CCC 720

Qy 1166 GCAATGATTACAGTGGCGAGATCTTGAACAACCTGCTGCTGATGAGTACCAAG-CC 1224
Db 721 GCAATGATTACAGTGGCGAGATCTTGAACAACCTGCTGCTGATGAGTACCAAG-CC 780

Qy 1225 ACAGGCCACCTTAGT-GCCCACTTCAGGAATATGTCCTGGAACGAATTAAGAGTCTAG 1282
Db 781 CCGGGCACCTTAGTGGCCCACTTCAGGAATATGTCCTGGAACGAATTAAGAGTCTAG 840

Qy 1283 ACCGTGCTGGGGCAGAGTCCGTGCACAGAGAAAAAATTTACAAT 1325
Db 841 GACCCGTGGGGGGCAAAAGTCCGGGGAACAGAGAAAAAT 883

RESULT 14
AL576279/c
LOCUS
DEFINITION AL576279 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1073YN20 3-PRIME, mRNA sequence.
ACCESSION AL576279
VERSION AL576279.3 GI:46249147
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:31314566.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
8244.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0D1073DGI0NPl&c=8244.r.
Location/Qualifiers
1. 991
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1073YN20"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
source
Query Match 31.9%; Score 753; DB 1; Length 991;
Best Local Similarity 94.3%; Pred. No. 1.5e-176;
Matches 766; Conservative 29; Mismatches 15; Indels 2; Gaps 2;

Qy 1479 GGTGCCATTGGCCGTGCTGACAAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACAT 1538
Db 987 GGGCCCTTGGSKGTGCTTCACAMAGTG-TKTGGCCACAGCTKTGTGA-GCGCTCAACAT 930

Db	493	GAGACTTGAATTACCTTATCTACGTGTTTCTCTGATCGGCCAAAGATGAAGTACTCCA	552
Qy	2042	AATACTACACACCAAGTTCCTCGGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTGA	2101
Db	553	AATACTACACACCAAGTTCCTCGGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTGA	612
Qy	2102	AGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGACGCACTCTGCAGATGCCGGGGCG	2161
Db	613	AGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGACGCACTCTGCAGATGCCGGGGCG	672
Qy	2162	GCAGCGCCACGTACATGGACCAAGGCCCTCCCGAGCTGTGTGTCCTCCAGGCTCACTATA	2221
Db	673	GCAGCGCCACGTACATGGACCAAGGCCCTCCCGAGCTGTGTGTCCTCCAGGCTCACTATA	731
Qy	2222	ACATGTACCCACAGAACCTTGACTCAGTCCTTGACACCGA	2261
Db	732	ACATGTACCCACAGAACCTTGACTCAGTCCTTGACACCGA	771

Search completed: May 17, 2005, 09:24:31
Job time : 11371 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2005, 18:30:18 ; Search time 1828 Seconds
(without alignments)

7655.508 Million cell updates/sec

Title: US-10-052-482-162

Perfect score: 2364

Sequence: 1 atggctgtggtgatacaagc.....tccgcagcgcacaatcgtga 2364

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_16Dec04.*

1: geneseqn1980s.*

2: geneseqn1990s.*

3: geneseqn2000s.*

4: geneseqn2001as.*

5: geneseqn2001bs.*

6: geneseqn2002as.*

7: geneseqn2002bs.*

8: geneseqn2003as.*

9: geneseqn2003bs.*

10: geneseqn2003cs.*

11: geneseqn2003ds.*

12: geneseqn2004as.*

13: geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2364	100.0	2364	9 ADA02656	ADA02656 Human STA
2	2364	100.0	2364	10 ADB72394	ADB72394 Human STA
3	2364	100.0	2364	10 ADE95904	ADE95904 Human STA
4	2364	100.0	2782	6 ABZ35457	ABZ35457 Human gen
5	2364	100.0	2782	6 ABK72299	ABK72299 Lymphoma
6	2364	100.0	2782	9 ADA02655	ADA02655 Human STA
7	2364	100.0	2782	10 ADB72393	ADB72393 Human STA
8	2364	100.0	2782	10 ADE95903	ADE95903 Human DNA
9	2356	99.7	2364	12 ADP82726	ADP82726 Human STA
10	2356	99.7	2716	6 ABZ35008	ABZ35008 Human gen
11	2356	99.7	2716	6 ABK84353	ABK84353 Human cDN
12	2313.2	97.9	2779	12 ADM79353	ADM79353 Human Lym
13	2170	91.8	2170	6 ABS51482	ABS51482 Human cDN
14	2073.2	87.7	2776	13 ACN40207	ACN40207 Tumour-as
15	1991.2	84.2	2671	10 ADB59436	ADB59436 Toxicity-
16	1991.2	84.2	2671	10 ADB53016	ADB53016 Primary r
17	1978.4	83.7	2940	6 ABK72273	ABK72273 Lymphoma
18	1976.8	83.6	2361	9 ADA02653	ADA02653 Mouse Sta
19	1976.8	83.6	2361	10 ADB72391	ADB72391 Mouse Sta
20	1976.8	83.6	2361	10 ADE95901	ADE95901 Mouse Sta

21	1976.8	83.6	2361	10 ACC79609	Acc79609 Mouse Sta
22	1976.8	83.6	2361	12 ADP82730	Adp82730 Mouse Sta
23	1976.8	83.6	2574	9 ADA02652	Ada02652 Mouse Sta
24	1976.8	83.6	2574	10 ADB72390	Adb72390 Mouse Sta
25	1976.8	83.6	2574	10 ADE95900	Ade95900 Mouse DNA
26	1972	83.4	2940	12 ADM79327	Adm79327 Mouse lym
27	1894.8	80.2	2385	2 AAT33590	Aat33590 Signal tr
28	1894.8	80.2	2385	12 ADP82724	Adp82724 Human STA
29	1894.8	80.2	3120	2 AAV71029	Aav71029 Green flo
30	1894.8	80.2	3138	2 AAV71041	Aav71041 Stat5-gre
31	1894.8	80.2	3714	5 AAA89231	Aaa89231 Human sig
32	1894.8	80.2	3714	11 ADI32144	Adi32144 Human cDN
33	1894.8	80.2	4279	6 ABK84328	Abk84328 Human cDN
34	1894.8	80.2	4279	10 ADH28822	Adh28822 Human chr
35	1894.8	80.2	4298	12 ADL83236	Adl83236 Human PRO
36	1823.6	77.1	2538	12 ADQ84604	Adq84604 Human tum
37	1770	74.9	2382	10 ACC79608	Acc79608 Mouse Sta
38	1770	74.9	2382	12 ADP82728	Adp82728 Mouse STA
39	1766.8	74.7	2382	12 ADP82732	Adp82732 STAT5A 1*
40	1766.8	74.7	2382	12 ADP82735	Adp82735 STAT5A 1*
41	1741	73.6	2818	2 AAT10554	Aat10554 Sheep mam
42	506	21.4	596	12 ADN00006	Adn00006 Human STA
43	428	18.1	446	12 ADN00007	Adn00007 Human STA
44	402.2	17.0	495	5 AAA89241	Aaa89241 Human STA
45	400.2	16.9	824	6 ABQ54442	Abq54442 Human ova

ALIGNMENTS

RESULT 1

ADA02656
ID ADA02656 standard; cDNA; 2364 BP.

XX ADA02656;

AC ADA02656;

DT 06-NOV-2003 (first entry)

DE Human STAT5B carcinoma associated coding sequence, SEQ ID NO:1174.

XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.

XX Homo sapiens.

OS Homo sapiens.
XX WO2003057146-A2.

PN WO2003057146-A2.

XX 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.

XX 26-DEC-2001; 2001US-00035832.

XX (SAGR-) SAGRES DISCOVERY.

PI Morris DW;

XX WPI; 2003-587068/55.

XX New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.

PS Claim 1; SEQ ID NO 1174; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic
XX acid sequences from mouse and human (ADA01482-ADA03094), and to
XX recombinant carcinoma associated proteins (CAP) encoded by them. The
XX invention also encompasses expression vectors and host cells comprising a
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically
XX binds to the protein, and a biochip comprising CA nucleic acid or
XX fragments thereof. The sequences of the invention were identified using
XX oncogenic retroviruses, which insert into the genome of the host organism

at random. Many of these do not carry transduced host oncogenes or pathogenic trans-acting viral genes, meaning that cancer incidence is a direct consequence of the effects of proviral integration into host protooncogenes. The CA nucleic acid sequences can be used to diagnose carcinoma (especially breast cancer, prostate cancer, lymphoma or leukemia) or a propensity to carcinoma by determination of the sequence of a CA gene, or by determination of CA gene expression in particular tissues. CA nucleic acids, proteins and antibodies are also useful as therapeutic agents and in screening and evaluating drug candidates. The present sequence represents a specifically claimed human CA nucleic acid sequence of the invention. Note: The complete sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at <http://wipo.int/pub/published/pct> sequences.

Sequence 2364 BP; 594 A; 636 C; 666 G; 468 T; 0 U; 0 Other; XX

Query Match	100.0%;	Score 2364;	DB 9;	Length 2364;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2364;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGSGCTGTGTGGATACAAAGCTCAGCAGCTCCAAAGSAGAAGCCCTTCATCAGATGCAAGCG	60	
Db	1	ATGSGCTGTGTGGATACAAAGCTCAGCAGCTCCAAAGSAGAAGCCCTTCATCAGATGCAAGCG	60	
QY	61	TTATATATGCCCAGCATTTTCCCATTTGAGGTGCGGCATTATTATTCCTCCAGTGGATTGAAAGC	120	
Db	61	TTATATATGCCCAGCATTTTCCCATTTGAGGTGCGGCATTATTATTCCTCCAGTGGATTGAAAGC	120	
QY	121	CAAGCATGGGACTCAGTAGATCTTGATAATCCACAGGAGAACAATTAAGGCCACCCAGCTC	180	
Db	121	CAAGCATGGGACTCAGTAGATCTTGATAATCCACAGGAGAACAATTAAGGCCACCCAGCTC	180	
QY	181	CTGAGAGGCCCTGTGTGACAGGAGCTGCAGAAAGAAGGCAGACACCAAGTGGGGGAAGATGGG	240	
Db	181	CTGAGAGGCCCTGTGTGACAGGAGCTGCAGAAAGAAGGCAGACACCAAGTGGGGGAAGATGGG	240	
QY	241	TTTTTTACTGAAGATCAAGCTGGGGCACTATGCGCACACAGCTCCAGAAACAAGTATGACCGC	300	
Db	241	TTTTTTACTGAAGATCAAGCTGGGGCACTATGCGCACACAGCTCCAGAAACAAGTATGACCGC	300	
QY	301	TGCCCCATGAGCTGGTCCGCTGCATCCGCCATATTTGTACATGAACACAGAGGTTGGTC	360	
Db	301	TGCCCCATGAGCTGGTCCGCTGCATCCGCCATATTTGTACATGAACACAGAGGTTGGTC	360	
QY	361	CSAGAAGCCAAACATGGTAGCTCTCCAGCTGGAGCCCTGCTGATGCCATGTCCCAGAAA	420	
Db	361	CSAGAAGCCAAACATGGTAGCTCTCCAGCTGGAGCCCTGCTGATGCCATGTCCCAGAAA	420	
QY	421	CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGGACTGGTCAACGACGACACAGAGAAT	480	
Db	421	CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGGACTGGTCAACGACGACACAGAGAAT	480	
QY	481	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACCAGGAGAGCGCTG	540	
Db	481	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACCAGGAGAGCGCTG	540	
QY	541	AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGCGCTCTGAGCCGG	600	
Db	541	AGGATCCAAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGCGCTCTGAGCCGG	600	
QY	601	GAGACGGCCCTCCAGCAGAGCAGGTTCTCTCGAGGCGCTGTTGACAGCTGAGGCACAG	660	
Db	601	GAGACGGCCCTCCAGCAGAGCAGGTTCTCTCGAGGCGCTGTTGACAGCTGAGGCACAG	660	
QY	661	ACACTGCAGCAGTAGTCCCGCTGGAGCTGCCCGAGAAGCACAGAGAAGACCTCTGCAGCTGCTG	720	
Db	661	ACACTGCAGCAGTAGTCCCGCTGGAGCTGCCCGAGAAGCACAGAGAAGACCTCTGCAGCTGCTG	720	
QY	721	CGGAGCAGCGAGACCATCTCTGGATGAAGAGCTGATCCAGTGGAAAGCGCGCAGCAG	780	
Db	721	CGGAGCAGCGAGACCATCTCTGGATGAAGAGCTGATCCAGTGGAAAGCGCGCAGCAG	780	

Db 1861 GACTCAGAAATGGCGCATCACCATTGCTTGGAAAGTTTGATCTCAGGAAGAATGTTT 1920
Qy 1921 TGGAAATCTGATGCTCTTTTACCACAGAGACTTCTCCATCAGTCCCTAGCGCGCGCTTG 1980
Db 1921 TGGAAATCTGATGCTCTTTTACCACAGAGACTTCTCCATCAGTCCCTAGCGCGCGCTTG 1980
Qy 1981 GGAGACTTGAATTAACCTTATCTAGCTGTTTCTCTGATCGGCCAAAGATGAAGTATACCTCC 2040
Db 1981 GGAGACTTGAATTAACCTTATCTAGCTGTTTCTCTGATCGGCCAAAGATGAAGTATACCTCC 2040
Qy 2041 AAATACTACACACAGTTCCTCGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2100
Db 2041 AAATACTACACACAGTTCCTCGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2100
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGGC 2160
Db 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGGC 2160
Qy 2161 GGCAAGCCACGATACATGACAGCCGCTCCCTCCAGCTGTGTCTCCAGGCTCACTAT 2220
Db 2161 GGCAAGCCACGATACATGACAGCCGCTCCCTCCAGCTGTGTCTCCAGGCTCACTAT 2220
Qy 2221 AACATGTACCCACAGAACCTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Db 2221 AACATGTACCCACAGAACCTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Qy 2281 GACCAATGGAGCTAGCGCGGCTGTGGAGAGCTCCTGGCGCGCAATGGACAGTCAG 2340
Db 2281 GACCAATGGAGCTAGCGCGGCTGTGGAGAGCTCCTGGCGCGCAATGGACAGTCAG 2340
Qy 2341 TGGATCCCGCACGACCAATCTGTA 2364
Db 2341 TGGATCCCGCACGACCAATCTGTA 2364

RESULT 2
ADB72394
ID ADB72394 standard; cDNA; 2364 BP.
XX ADB72394;
AC ADB72394;
XX
DT 04-DEC-2003 (first entry)
XX Human STAT5B cDNA.
XX human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX Homo sapiens.
XX WO2003008583-A2.
XX 30-JAN-2003.
XX
XX 26-DEC-2001; 2001WO-US051291.
XX
XX 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
PA Morris DW, Engelhard EK;
XX WPI; 2003-239337/23.
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
XX Claim 1; SEQ ID NO 222; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytosatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human cDNA of the invention.
XX
SQ Sequence 2364 BP; 594 A; 636 C; 666 G; 468 T; 0 U; 0 Other;

Query Match 100.0%; Score 2364; DB 10; Length 2364;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCTCTGTGGATACAAAGCTCAGCAGCTCAAGGAGAAGCCCTTCATCAGATGCAAGCG 60
Db 1 ATGCTCTGTGGATACAAAGCTCAGCAGCTCAAGGAGAAGCCCTTCATCAGATGCAAGCG 60

Qy 61 TTATATGCCAGCAATTTTCCATTGAGTGGCGCATTTATTTATCCAGTGGATTTGAAGC 120
Db 61 TTATATGCCAGCAATTTTCCATTGAGTGGCGCATTTATTTATCCAGTGGATTTGAAGC 120

Qy 121 CAAGCATGGGACTCAGTAGATCTTGAATATCCAGGAGAACATTAAGGCCACCCAGCTC 180
Db 121 CAAGCATGGGACTCAGTAGATCTTGAATATCCAGGAGAACATTAAGGCCACCCAGCTC 180

Qy 181 CTGAGGGGCTGTGTGAGGAGCTGCAGAGAAGGACAGCACAGGTGGGGAGATGGG 240
Db 181 CTGAGGGGCTGTGTGAGGAGCTGCAGAGAAGGACAGCACAGGTGGGGAGATGGG 240

Qy 241 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAACACATGACCGC 300
Db 241 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAACACATGACCGC 300

Qy 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCGCATATTTGTACAATGAACAGAGTTGGTC 360
Db 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCGCATATTTGTACAATGAACAGAGTTGGTC 360

Qy 361 CGAAGCCCAACATGATGCTCCAGCTGGAGCCCTTCTGATGCCATGCTCCAGAAA 420
Db 361 CGAAGCCCAACATGATGCTCCAGCTGGAGCCCTTCTGATGCCATGCTCCAGAAA 420

Qy 421 CACCTCCAGATCAACAGAGCTTTGAGGAGCTGGGACTGGTCCAGGAGACACAGAGAT 480
Db 421 CACCTCCAGATCAACAGAGCTTTGAGGAGCTGGGACTGGTCCAGGAGACACAGAGAT 480

Qy 481 GAGTAAAAAAGCTGCAGCAGACTCAGAGGTACTTTCATCATCCAGTACAGGAGCCCTG 540
Db 481 GAGTAAAAAAGCTGCAGCAGACTCAGAGGTACTTTCATCATCCAGTACAGGAGCCCTG 540

Qy 541 AGGATCCCAAGCTCAGTTTGGCCCGCTGGCCCGCTGAGCCCGCCAGGCGCTGAGCCGG 600
Db 541 AGGATCCCAAGCTCAGTTTGGCCCGCTGGCCCGCTGAGCCCGCCAGGCGCTGAGCCGG 600

Qy 601 GAGACGGCCCTCCAGCAGAGAGCAGTGTCTCTGGAGGCTGGTTCAGCGCTGAGGCACAG 660
Db 601 GAGACGGCCCTCCAGCAGAGAGCAGTGTCTCTGGAGGCTGGTTCAGCGCTGAGGCACAG 660

Qy 661 ACCTGCAGCAGTACCGCGTGGAGCTGCCGAGAGACACAGAGAGCCCTGAGCTGCTG 720
Db 661 ACCTGCAGCAGTACCGCGTGGAGCTGCCGAGAGACACAGAGAGCCCTGAGCTGCTG 720

Qy 721 CGGAAGCAGCAGACCATCTCTGGATGACGAGCTGATCCAGTGAAGCGCGCGCAGCAG 780
Db 721 CGGAAGCAGCAGACCATCTCTGGATGACGAGCTGATCCAGTGAAGCGCGCGCAGCAG 780

Qy 781 CTGSCCGGGAACGGCGGGCCCCCGAGCGAGCTGGACGCTGTACAGTCTCTGGTGTGAG 840
Db 781 CTGSCCGGGAACGGCGGGCCCCCGAGCGAGCTGGACGCTGTACAGTCTCTGGTGTGAG 840

Qy 841 AAGTTGGCGGAGATCATCTGCGAGAACCGGCGAGCAGATCCGACGGGTGACACCTCTGC 900
Db 841 AAGTTGGCGGAGATCATCTGCGAGAACCGGCGAGCAGATCCGACGGGTGACACCTCTGC 900

841 AAGTTGGGAGATCATCTGCGAAGACCGGACGAGATCCGACGGGCTGACCACTCTGC 900
901 CAGCAGCTGCCATCCCGGCCCAAGTGGAGAGATGCTGGCGAGGTCAACGCCACCATC 960
901 CAGCAGCTGCCATCCCGGCCCAAGTGGAGAGATGCTGGCGAGGTCAACGCCACCATC 960
961 ACGACATTAATCTCAGCCCTGGTACCGACGACGTTTCATCTGAGAAGCAGCCTCCTCAG 1020
961 ACGACATTAATCTCAGCCCTGGTACCGACGACGTTTCATCTGAGAAGCAGCCTCCTCAG 1020
1021 GTCTGGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCCTGTGTTGGGGCGGAAGCTG 1080
1021 GTCTGGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCCTGTGTTGGGGCGGAAGCTG 1080
1081 AACGTGCACATGAACCCCCCCCAGGTGAAGGCCACCATCATCATGAGCAGCAGGCCAAG 1140
1081 AACGTGCACATGAACCCCCCCCAGGTGAAGGCCACCATCATCATGAGCAGCAGGCCAAG 1140
1141 TCTCTGCTCAAGAACGAGAACCCGCAATGATTTACAGTGGCGAGATCTTGAACAACTGC 1200
1141 TCTCTGCTCAAGAACGAGAACCCGCAATGATTTACAGTGGCGAGATCTTGAACAACTGC 1200
1201 TGCGTCAATGGAGTACCAAGGCCACAGGCCACCTTAGTGCCCACTTCAGGAATATGTC 1260
1201 TGCGTCAATGGAGTACCAAGGCCACAGGCCACCTTAGTGCCCACTTCAGGAATATGTC 1260
1261 CTGAAACGAAATTAAGAGGTCAAGCCGTGTTGGGGCAGAGTGGTGAAGAGAAAAATTT 1320
1261 CTGAAACGAAATTAAGAGGTCAAGCCGTGTTGGGGCAGAGTGGTGAAGAGAAAAATTT 1320
1321 ACAATCTGTTGAATCCAGTTCAGTGTGGTGGAAATGAGCTGGTTTTCAAGTCAAG 1380
1321 ACAATCTGTTGAATCCAGTTCAGTGTGGTGGAAATGAGCTGGTTTTCAAGTCAAG 1380
1381 ACCGTGCTCCCTGCGAGTGGTGTGATCTTATGCGCAGCAGGACAAATGCGACGGCC 1440
1381 ACCGTGCTCCCTGCGAGTGGTGTGATCTTATGCGCAGCAGGACAAATGCGACGGCC 1440
1441 ACTGTTCTTGGGACAAATGCTTTTGCAGAGCCTGCGAGGGTGCCATTTGCGTGCCTGAC 1500
1441 ACTGTTCTTGGGACAAATGCTTTTGCAGAGCCTGCGAGGGTGCCATTTGCGTGCCTGAC 1500
1501 AAAGTCTGTGGCCACAGCTGTGTGAGCGCTCAACATGAATTCAGAGCGCGAAGTGCAG 1560
1501 AAAGTCTGTGGCCACAGCTGTGTGAGCGCTCAACATGAATTCAGAGCGCGAAGTGCAG 1560
1561 AGCAACCGGGCCTGACCAAGGAGAACTCTGTGTTCTGGCGCAGAAACTGTTCAACAAC 1620
1561 AGCAACCGGGCCTGACCAAGGAGAACTCTGTGTTCTGGCGCAGAAACTGTTCAACAAC 1620
1621 AGCAGACGCCACCTGGAGGACTACAGTGGCCTGTCTGTGCTCTGCTGCCAGTTCAACAGG 1680
1621 AGCAGACGCCACCTGGAGGACTACAGTGGCCTGTCTGTGCTCTGCTGCCAGTTCAACAGG 1680
1681 GAGAAATTACAGACCGGAATTACACTTTCGCGCAATGGTTTGACGGTGTGATGGAAGTG 1740
1681 GAGAAATTACAGACCGGAATTACACTTTCGCGCAATGGTTTGACGGTGTGATGGAAGTG 1740
1741 TTAATAAACAATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTTTGTAAACAAG 1800
1741 TTAATAAACAATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTTTGTAAACAAG 1800
1801 CAACAGGCCCATGACCTACTGATTAACAAGCCAGATGGGACCTTCCTCCTGAGATTTCAGT 1860
1801 CAACAGGCCCATGACCTACTGATTAACAAGCCAGATGGGACCTTCCTCCTGAGATTTCAGT 1860
1861 GACTCAGAAATGGGGCATCACCATTGCTTGGAGTTTGAATTCAGGAAGAATGTTT 1920
1861 GACTCAGAAATGGGGCATCACCATTGCTTGGAGTTTGAATTCAGGAAGAATGTTT 1920
1921 TGGAAATCTGATGCTTTTACCACAGAGACTTTCCTCAATCAGTCCCTAGCGCAGCGCTTG 1980
1921 TGGAAATCTGATGCTTTTACCACAGAGACTTTCCTCAATCAGTCCCTAGCGCAGCGCTTG 1980

Qy 1981 GGAGACTTGAATTAACCTTATCTAGCTGTTTCTGATCGGCCAAAAGATGAAGTATACTCC 2040
Db 1981 GGAGACTTGAATTAACCTTATCTAGCTGTTTCTGATCGGCCAAAAGATGAAGTATACTCC 2040
Qy 2041 AAATACTACACACAGTTCCCTCGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2100
Db 2041 AAATACTACACACAGTTCCCTCGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2100
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTCTGTTGAAACGATCTGCAGATGCCGGGGC 2160
Db 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTCTGTTGAAACGATCTGCAGATGCCGGGGC 2160
Qy 2161 GGCAGCGCCACGTAATGAGACCGGCCCTCCCGAGCTGTGTCGCCAGGCTCACTAT 2220
Db 2161 GGCAGCGCCACGTAATGAGACCGGCCCTCCCGAGCTGTGTCGCCAGGCTCACTAT 2220
Qy 2221 AACATGTACCCACAGAACCCCTGACTCAGTCTGTTGACACCGATGGGAGCTTCGATCTGGAG 2280
Db 2221 AACATGTACCCACAGAACCCCTGACTCAGTCTGTTGACACCGATGGGAGCTTCGATCTGGAG 2280
Qy 2281 GACACAATGACAGCTAGCGCGCGCTGTGGAGGAGCTCCTGGCGCGGCCAAATGGAAGTCAAG 2340
Db 2281 GACACAATGACAGCTAGCGCGCGCTGTGGAGGAGCTCCTGGCGCGGCCAAATGGAAGTCAAG 2340
Qy 2341 TGGATCCCGCAGCAGCAATCGTGA 2364
Db 2341 TGGATCCCGCAGCAGCAATCGTGA 2364

RESULT 3
ADE95904

ID ADE95904 standard; DNA; 2364 BP.

XX AC ADE95904;

XX DT 12-FEB-2004 (first entry)

XX Human STAT5B gene coding sequence.

XX DE cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
XX KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; STAT5B.

XX OS Homo sapiens.

XX PN WO2003039484-A2.

XX PD 15-MAY-2003.

XX PF 08-NOV-2002; 2002WO-US036071.

XX PR 08-NOV-2001; 2001US-00052482.

XX XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX XX WPI; 2003-441462/41.

XX DR New carcinoma associated nucleic acids and proteins, useful for screening
XX PT drug candidates, or for diagnosing and treating carcinomas, e.g.
XX PT lymphoma, breast cancer, prostate cancer or leukemia.

XX PS Claim 1; SEQ ID NO 162; 793pp; English.

XX CC This invention relates to novel recombinant nucleic acids for use in
XX CC diagnosis and treatment of cancer, especially carcinomas, as well as the
XX CC use of compositions in screening methods. The compositions of the
XX CC invention may have cytostatic activity whilst the disclosed sequences may
XX CC be useful for gene therapy. The carcinoma associated nucleic acids and
XX CC proteins are useful for diagnosing and treating carcinomas, for example
XX CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
XX CC drug candidates or bioactive agents capable of binding to, or modulating

CC the activity of, a carcinoma associated protein. The present sequence is
CC the coding DNA sequence of the human STAT5B gene which is a carcinoma
CC associated gene of the invention.

XX
SQ Sequence 2364 BP; 594 A; 636 C; 666 G; 468 T; 0 U; 0 Other;
Query Match 100.0%; Score 2364; DB 10; Length 2364;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGCTGTGGATACAGCTCAGAGCTCAAGAGAGAGCCCTTCATCATGCAAGCG 60
Db 1 ATGGCTGTGGATACAGCTCAGAGCTCAAGAGAGAGCCCTTCATCATGCAAGCG 60
Qy 61 TTATATGCCAGCATTTTCCCATTCAGGTGGGCAATTTATCCAGTGAATGAAGC 120
Db 61 TTATATGCCAGCATTTTCCCATTCAGGTGGGCAATTTATCCAGTGAATGAAGC 120
Qy 121 CAAGCATGGGACTCAGTAGATCTTGAATATCCAGAGAGACATTTAAGGCCACCCAGCTC 180
Db 121 CAAGCATGGGACTCAGTAGATCTTGAATATCCAGAGAGACATTTAAGGCCACCCAGCTC 180
Qy 181 CTGAGGCGCTGGTGCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 181 CTGAGGCGCTGGTGCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Qy 241 TTTTACTGAAGATCAAGCTGGGCACTATGCCACAGCTCCAGAACACGTATGACCGC 300
Db 241 TTTTACTGAAGATCAAGCTGGGCACTATGCCACAGCTCCAGAACACGTATGACCGC 300
Qy 301 TGCSCCATGGAGCTGGTCCGCTGCATCCGCGCATATATTGTACAAATGAACAGAGTTGGTC 360
Db 301 TGCSCCATGGAGCTGGTCCGCTGCATCCGCGCATATATTGTACAAATGAACAGAGTTGGTC 360
Qy 361 CGAAGAGCAACAAATGGTAGCTCTCCAGCTGGAAGCTTGTGTGATGCCAAGAAA 420
Db 361 CGAAGAGCAACAAATGGTAGCTCTCCAGCTGGAAGCTTGTGTGATGCCAAGAAA 420
Qy 421 CACCTCCAGATCAACAGAGCTTTGAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 480
Db 421 CACCTCCAGATCAACAGAGCTTTGAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAG 480
Qy 481 GAGTTAAAAAGCTGCAGCAGACTCAGAGTACTTTCATCATCCAGTACAGAGAGCCTG 540
Db 481 GAGTTAAAAAGCTGCAGCAGACTCAGAGTACTTTCATCATCCAGTACAGAGAGCCTG 540
Qy 541 AGGATCCAAAGCTCAGTTTGGCCGCTGCCAGCTGAGCCCCCAGAGAGCTGTGAGCGG 600
Db 541 AGGATCCAAAGCTCAGTTTGGCCGCTGCCAGCTGAGCCCCCAGAGAGCTGTGAGCGG 600
Qy 601 GAGAGGCGCTCCAGCAGAGAGAGTGTCTTGGAGGCTGGTGGAGGCTGGAGGCTGGAG 660
Db 601 GAGAGGCGCTCCAGCAGAGAGAGTGTCTTGGAGGCTGGTGGAGGCTGGAGGCTGGAG 660
Qy 661 ACATGTCAGAGTACCGGTGGAGCTGCCAGAGAGACACAGAGAGCCTTCAGCTGCTG 720
Db 661 ACATGTCAGAGTACCGGTGGAGCTGCCAGAGAGACACAGAGAGCCTTCAGCTGCTG 720
Qy 721 CGGAAGCAGCAGACCATCATCTGGATGACAGCTGATCCAGTGGAAAGCGGCGAGCAG 780
Db 721 CGGAAGCAGCAGACCATCATCTGGATGACAGCTGATCCAGTGGAAAGCGGCGAGCAG 780
Qy 781 CTGGCGGGAAACGGGCGGCGCGCGAGGCGAGCTTGGAGCTGCTACAGTCTGTGTGAG 840
Db 781 CTGGCGGGAAACGGGCGGCGCGCGAGGCGAGCTTGGAGCTGCTACAGTCTGTGTGAG 840
Qy 841 AAGTTGGCGGAGATCATCTGCAGAAACCGGAGAGATCCGAGGCTGAGCACTCTGC 900
Db 841 AAGTTGGCGGAGATCATCTGCAGAAACCGGAGAGATCCGAGGCTGAGCACTCTGC 900
Qy 901 CAGCAGCTGCCCATCCCCGGCGCGAGTGGAGAGATGCTGGCGGAGGTCAAGCGCACCATC 960
Db 901 CAGCAGCTGCCCATCCCCGGCGCGAGTGGAGAGATGCTGGCGGAGGTCAAGCGCACCATC 960

Qy 961 ACGGACATATCTCAGCCCTGGTGACGACGACGTTTCATTCATTGAGAAAGACGCTCCTCAG 1020
Db 961 ACGGACATATCTCAGCCCTGGTGACGACGACGTTTCATTCATTGAGAAAGACGCTCCTCAG 1020
Qy 1021 GTCTGAAGACCCAGACCAAGTTTGCAGCCTCTGTGGCCCTGTGGTGGGCGGAAAGCTG 1080
Db 1021 GTCTGAAGACCCAGACCAAGTTTGCAGCCTCTGTGGCCCTGTGGTGGGCGGAAAGCTG 1080
Qy 1081 AACGTGCACATGAACCCCGCCAGGTGAAGGCCACCATCATCATGTGAGCAGAGCCCAAG 1140
Db 1081 AACGTGCACATGAACCCCGCCAGGTGAAGGCCACCATCATCATGTGAGCAGAGCCCAAG 1140
Qy 1141 TCTCTGTCTCAAGAACGAGAACACCCCGCNAATGATACAGTGGCGAGATCTTCAAAACAATGC 1200
Db 1141 TCTCTGTCTCAAGAACGAGAACACCCCGCNAATGATACAGTGGCGAGATCTTCAAAACAATGC 1200
Qy 1201 TGCCTCATGTGAGTACCAAGCCACAGGCACTTATAGTGCCTCTTCAGGAAATATGTCC 1260
Db 1201 TGCCTCATGTGAGTACCAAGCCACAGGCACTTATAGTGCCTCTTCAGGAAATATGTCC 1260
Qy 1261 CTGAAACGAATTAAGAGGTGAGACCGTCTGGGCGAGAGTGGTGACAGAGAAATTT 1320
Db 1261 CTGAAACGAATTAAGAGGTGAGACCGTCTGGGCGAGAGTGGTGACAGAGAAATTT 1320
Qy 1321 ACAATCTGTTTGAATCCCAAGTTCAGTGTGGTGAATAGCTGGTGTTCAGAGTCAAG 1380
Db 1321 ACAATCTGTTTGAATCCCAAGTTCAGTGTGGTGAATAGCTGGTGTTCAGAGTCAAG 1380
Qy 1381 ACCCTGTCCCTGCGAGTGGTGTATCTTCATCGGAGCCAGGACAAACAATGCCAGCGCC 1440
Db 1381 ACCCTGTCCCTGCGAGTGGTGTATCTTCATCGGAGCCAGGACAAACAATGCCAGCGCC 1440
Qy 1441 ACTGTTCTGGGACATGCTTTTGCAGAGCTGGCAGGCTGCCATTTGCCGTGCTGAC 1500
Db 1441 ACTGTTCTGGGACATGCTTTTGCAGAGCTGGCAGGCTGCCATTTGCCGTGCTGAC 1500
Qy 1501 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAATTAAGAGCCGAGTGCAG 1560
Db 1501 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAATTAAGAGCCGAGTGCAG 1560
Qy 1561 AGCAACCGGGCTGACCAAGAGAACTCGTGTCTTGGCGCAGAACTGTTCACAAAC 1620
Db 1561 AGCAACCGGGCTGACCAAGAGAACTCGTGTCTTGGCGCAGAACTGTTCACAAAC 1620
Qy 1621 AGCAGCAGCCTTGGAGGACTACAGTGGCTGTGTGTCTGTCTGCTGCTGCTGCTGCTG 1680
Db 1621 AGCAGCAGCCTTGGAGGACTACAGTGGCTGTGTGTCTGTCTGCTGCTGCTGCTGCTG 1680
Qy 1681 GAGAAATTTACAGGACGGAATTTACATTTCTTGGCAATGGTTTGGCGTGTGATGGAAGTG 1740
Db 1681 GAGAAATTTACAGGACGGAATTTACATTTCTTGGCAATGGTTTGGCGTGTGATGGAAGTG 1740
Qy 1741 TTAATAAAACATCTCAAGCCCTCATTTGGAATGATGGGCAATTTTGGGTTTGTAAACAAG 1800
Db 1741 TTAATAAAACATCTCAAGCCCTCATTTGGAATGATGGGCAATTTTGGGTTTGTAAACAAG 1800
Qy 1801 CAAAGCGCCCATGACCTACTGATTAACAGCCAGATGGGACCTTCTCTCCTGAGATTCAGT 1860
Db 1801 CAAAGCGCCCATGACCTACTGATTAACAGCCAGATGGGACCTTCTCTCCTGAGATTCAGT 1860
Qy 1861 GACTCAGAAATTTGGCGCATCACCATTGCTTGGAGTTTGAATCTTCAGGAAGAAATGTTT 1920
Db 1861 GACTCAGAAATTTGGCGCATCACCATTGCTTGGAGTTTGAATCTTCAGGAAGAAATGTTT 1920
Qy 1921 TGGAAATCTGATGCTTTTACCACAGAGACTTCTTCATCAGGTCCCTAGCGGACCGCTTG 1980
Db 1921 TGGAAATCTGATGCTTTTACCACAGAGACTTCTTCATCAGGTCCCTAGCGGACCGCTTG 1980
Qy 1981 GGAGACTTGAATTAACCTTATCTACGTGTTCTGATCGGCCAAAGATGAAGTATATCTCC 2040
Db 1981 GGAGACTTGAATTAACCTTATCTACGTGTTCTGATCGGCCAAAGATGAAGTATATCTCC 2040

QY 2041 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTGTGATGATACGTG 2100
DB |||||||
QY 2041 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTGTGATGATACGTG 2100
DB |||||||
QY 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGGC 2160
DB |||||||
QY 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGGC 2160
DB |||||||
QY 2161 GGCAGCGCAGTACATGACAGAGCCCTCCCTCCAGCTGTGTGCCAGGCTCACTAT 2220
DB |||||||
QY 2161 GGCAGCGCCAGTACATGACAGAGCCCTCCCTCCAGCTGTGTGCCAGGCTCACTAT 2220
DB |||||||
QY 2221 AACATGTACCCACAGAACCTGACTCAGTCTTTCACACCGATGGGGACTTCGATCTCGAG 2280
DB |||||||
QY 2221 AACATGTACCCACAGAACCTGACTCAGTCTTTCACACCGATGGGGACTTCGATCTCGAG 2280
DB |||||||
QY 2281 GACACAAATGGACGTAGCGCGCGCTGTGGAGGAGTCTCTGGCGCCGCAATGGACAGTCAG 2340
DB |||||||
QY 2281 GACACAAATGGACGTAGCGCGCGCTGTGGAGGAGTCTCTGGCGCCGCAATGGACAGTCAG 2340
DB |||||||
QY 2341 TGGATCCCGCAGCAGCAATCGTGA 2364
DB |||||||
QY 2341 TGGATCCCGCAGCAGCAATCGTGA 2364
DB |||||||

RESULT 4

ABZ35457
ID ABZ35457 standard; cDNA; 2782 BP.
AC ABZ35457;
XX
DT 05-FEB-2003 (first entry)
XX
DE Human gene expression profile polynucleotide SEQ ID NO 568.
XX
KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
KW gene expression; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200274979-A2.
XX
PD 26-SEP-2002.
XX
PF 20-MAR-2002; 2002WO-US008456.
XX
PR 20-MAR-2001; 2001US-0276947P.
XX
PA (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
XX
PI Wan J, Wang Y;
XX
XX WPI; 2002-740862/80.
XX
PT New gene expression profile generated from primary, endothelial,
PT epithelial, and muscle cell types, useful for identifying disease
PT pathologies involving alterations of gene expression, e.g. cancer.
XX
XX Disclosure; Page 704-705; 850pp; English.
XX
CC The invention relates to a gene expression profile comprising one or more
CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type
CC is a coronary artery endothelium, umbilical artery or vein endothelium,
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery
CC endothelium, myometrium microvascular endothelium, keratinocyte
CC epithelium, bronchial epithelium, mammary epithelium, prostate
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,
CC small airway epithelium, renal epithelium, umbilical artery smooth
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
CC

CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
CC osteoblasts or prostate stromal cell. The gene expression profile is used
CC for determining the level of RNA expression for a sample, determining the
CC phenotype of a cell and distinguishing cell types. The gene or a protein
CC expression profile is useful in identifying disease pathologies involving
CC alterations of gene expression. The assessment of expression profiles may
CC provide meaningful information with respect to tumour type and stage,
CC treatment methods, and prognosis. The gene or protein expression profile
CC may also be used for creating microarrays. The microarray is useful for
CC genetic and physical mapping of genomes, DNA sequencing, genetic or
CC medical diagnosis, genotyping of organisms, confirming cell or tissue
CC identifications and in identifying promising antibiotics, antiviral or
XX antifungal agents
SQ Sequence 2782 BP; 669 A; 739 C; 801 G; 573 T; 0 U; 0 Other;
Query Match 100.0%; Score 2364; DB 6; Length 2782;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTGTGTGATACAGCTCAGAGCTCCAGGAGAGCCCTTCATCATGATGCAAGCG 60
DB |||||||
QY 61 TTATATGGCCAGCATTTTCCCATTTGAGTGGCGCATTTATTTATCCAGTGGATTGAAGC 120
DB |||||||
QY 207 TTATATGGCCAGCATTTTCCCATTTGAGTGGCGCATTTATTTATCCAGTGGATTGAAGC 266
DB |||||||
QY 121 CAAGCATGGGACTCAGTAGATCTTTGATAATCCAGAGAACATTTAAGGCCACCCAGCTC 180
DB |||||||
QY 267 CAAGCATGGGACTCAGTAGATCTTTGATAATCCAGAGAACATTTAAGGCCACCCAGCTC 326
DB |||||||
QY 181 CTGAGGCGCTGTGTCAGGAGCTGTCAGAGAGAGAGACACAGAGTGGGGAGATGGG 240
DB |||||||
QY 327 CTGAGGCGCTGTGTCAGGAGCTGTCAGAGAGAGAGACACAGAGTGGGGAGATGGG 386
DB |||||||
QY 241 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC 300
DB |||||||
QY 387 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC 446
DB |||||||
QY 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCGCATATATTGTACAATGAACAGAGTTGGTC 360
DB |||||||
QY 447 TGCCCCATGGAGCTGGTCCGCTGCATCCGCGCATATATTGTACAATGAACAGAGTTGGTC 506
DB |||||||
QY 361 CGAGAGCCAAATATGCTAGCTCTCCAGCTGGAGAGCTTCTGATGTCATGTCACAGAAA 420
DB |||||||
QY 507 CGAGAGCCAAATATGCTAGCTCTCCAGCTGGAGAGCTTCTGATGTCATGTCACAGAAA 566
DB |||||||
QY 421 CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGCGACTGGTCCAGCAGGACACAGAGAT 480
DB |||||||
QY 567 CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGCGACTGGTTCAGCAGGACACAGAGAT 626
DB |||||||
QY 481 GAGTTAAAAAGCTGCAGCAGAGCTCAGGAGTACTTTCATCATCCAGTACCAGGAGAGCTG 540
DB |||||||
QY 627 GAGTTAAAAAGCTGCAGCAGAGCTCAGGAGTACTTTCATCATCCAGTACCAGGAGAGCTG 686
DB |||||||
QY 541 AGGATCCAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGAGCTCTGAGCGG 600
DB |||||||
QY 687 AGGATCCAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGAGCTCTGAGCGG 746
DB |||||||
QY 601 GAGACGCGCTCCAGCAGAGAGAGTGTCTCTGAGGCGCTGGTTGACGCGTGAGGACAG 660
DB |||||||
QY 747 GAGACGCGCTCCAGCAGAGAGAGTGTCTCTGAGGCGCTGGTTGACGCGTGAGGACAG 806
DB |||||||
QY 661 ACATCTGACAGCAGTACCGCTGGAGCTGCCCGAGAGACACAGAGACCTCTGACGCTGTG 720
DB |||||||
QY 807 ACATCTGACAGCAGTACCGCTGGAGCTGCCCGAGAGACACAGAGACCTCTGACGCTGTG 866
DB |||||||
QY 721 CGGAGCAGCAGACCATCATCTCGGATGACAGCTGATCCAGTGGAGAGCGCGGACAGCAG 780
DB |||||||
QY 867 CGGAGCAGCAGACCATCATCTCTGGATGACAGCTGATCCAGTGGAGAGCGCGGACAGCAG 926
DB |||||||
QY 781 CTGSCCGGGAACGCGCGGCGCCCCCGGAGGCGAGCTGGACGCTGTGCTACAGTCTCTGTGTGAG 840
DB |||||||

CC The invention describes a recombinant protein (I) from a lymphoma
CC associated protein (LAP) sequence of Plk3r1, GNAS (including
CC XI₁ (alpha_s), and NESP55), JAK1, Neurogranin or Nfr2 proteins. (I) is
CC useful for screening for a bioactive agent capable of binding to an LA
CC protein (LAP) which is encoded by a polynucleotide (II) and (II) is
CC useful for evaluating the effect of a candidate lymphoma drug in a
CC patient. (I) and (II) is also useful for diagnosing lymphoma involving
CC determining the expression of one or more (II), or (I) encoded by (II) in
CC a first tissue type of a first individual and comparing the expression of
CC the gene(s) from a second normal tissue type from the first individual or
CC a second unaffected individual, where a difference in the expression
CC indicates that the first individual has lymphoma. (II) is useful in the
CC diagnosis and treatment of lymphoma and leukaemia. A host cell comprising
CC (II) is useful for screening drug candidates which involves providing the
CC host cell that expresses LA gene, adding a drug candidate to the cell,
CC and determining the effect of the drug candidate on the expression of the
CC LA gene. LA genes are also administered as DNA vaccines, such that
CC expression of the polypeptide encoded by the DNA vaccine, cytotoxic T-
CC cells and antibodies are induced which recognise and destroy or eliminate
CC cells expressing LA proteins. The genes are also for generating animal
CC models of lymphoma. This sequence represents a lymphoma associated
CC polynucleotide described in the invention
XX

SQ Sequence 2782 BP; 669 A; 739 C; 801 G; 573 T; 0 U; 0 Other;

Query Match 100.0%; Score 2364; DB 6; Length 2782;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTGTGTGGATACAAAGCTCACAGCTCCAGGAGAACCCCTTCATCATGATCGAAGCG 60
DB 147 ATGCTGTGTGGATACAAAGCTCACAGCTCCAGGAGAACCCCTTCATCATGATCGAAGCG 206
QY 61 TTATATGCCAGCAATTTCCCAATGAGGTGGCGCATTTATTTATCCAGTGGATTTGAAGC 120
DB 207 TTATATGCCAGCAATTTCCCAATGAGGTGGCGCATTTATTTATCCAGTGGATTTGAAGC 266
QY 121 CAAGCATGGGACTCAGTAGATCTTGATTAATCCAGAGAGAACATTAAAGGCCACCCAGCTC 180
DB 267 CAAGCATGGGACTCAGTAGATCTTGATTAATCCAGAGAGAACATTAAAGGCCACCCAGCTC 326
QY 181 CTGAGGCGCTGTGTGAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 327 CTGAGGCGCTGTGTGAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 386
QY 241 TTTTACTGAAGATCAAGCTGGGCACTATGSCACACAGCTCCAGAACACGTATGACCGC 300
DB 387 TTTTACTGAAGATCAAGCTGGGCACTATGSCACACAGCTCCAGAACACGTATGACCGC 446
QY 301 TGCCCCATGGAGCTGGTCCGCTGATCGGCCATATATTGTACATGAACAGAGATTGGTC 360
DB 447 TGCCCCATGGAGCTGGTCCGCTGATCGGCCATATATTGTACATGAACAGAGATTGGTC 506
QY 361 CGAAGACCAACATGTGTAGCTCTCCAGCTGGAGCTTGTGATGCCATGCTCCAGAAA 420
DB 507 CGAAGACCAACATGTGTAGCTCTCCAGCTGGAGCTTGTGATGCCATGCTCCAGAAA 566
QY 421 CACCTCCAGATCAACACAGACGTTTGGAGAGCTGGAGCTGCTCAGCAGGACACAGAGAA 480
DB 567 CACCTCCAGATCAACACAGACGTTTGGAGAGCTGGAGCTGCTCAGCAGGACACAGAGAA 626
QY 481 GAGTTAAAGCTGACAGAGACTCAGAGTACTTTTCATCATCATGATACAGAGAGAGCTG 540
DB 627 GAGTTAAAGCTGACAGAGACTCAGAGTACTTTTCATCATCATGATACAGAGAGAGCTG 686
QY 541 AGGATCAAGCTCAGTTTGGCCCGCTGCCAGCTGAGCCGCCAGAGAGCTGCTGAGCCGG 600
DB 687 AGGATCAAGCTCAGTTTGGCCCGCTGCCAGCTGAGCCGCCAGAGAGCTGCTGAGCCGG 746
QY 601 GAGACGCCCTCCAGCAGAGAGAGAGTGTCTCTGGAGGCTGTGGTTGACCGCTGAGGACAG 660
DB 747 GAGACGCCCTCCAGCAGAGAGAGAGTGTCTCTGGAGGCTGTGGTTGACCGCTGAGGACAG 806

QY 661 AACTGACAGTAGTACCGCTGGAGCTGCCGAGAGACACAGAGACCCCTGCAGCTGCTG 720
DB 807 AACTGACAGTAGTACCGCTGGAGCTGCCGAGAGACACAGAGACCCCTGCAGCTGCTG 866
QY 721 CGGAAGCAGCAGACCATCATCTCTGGATGACAGCTGATCCAGTGAAGCGCGCAGCAG 780
DB 867 CGGAGCAGCAGACCATCATCTCTGGATGACAGCTGATCCAGTGAAGCGCGCAGCAG 926
QY 781 CTGCGCGGGAACCGCGGCGCCCCCGAGGGCAGCTGACGTGTACAGTCTCTGGTGTGAG 840
DB 927 CTGCGCGGGAACCGCGGCGCCCCCGAGGGCAGCTGACGTGTACAGTCTCTGGTGTGAG 986
QY 841 AAGTTGGCGGAGATCATCTGCGAGAACCGCGCAGCAGATCCGACGGGTGACACCTTGC 900
DB 987 AAGTTGGCGGAGATCATCTGCGAGAACCGCGCAGCAGATCCGACGGGTGACACCTTGC 1046
QY 901 CAGCAGCTGCCCATCCCGCGCCAGTGTGAGAGATGCTGGCCGAGGTCAACGCCACCATC 960
DB 1047 CAGCAGCTGCCCATCCCGCGCCAGTGTGAGAGATGCTGGCCGAGGTCAACGCCACCATC 1106
QY 961 ACGGACATTTATCTCAGCCCTGGTGAACAGCACGTTTATCATTTAGAGAGACGCTCTCAG 1020
DB 1107 ACGGACATTTATCTCAGCCCTGGTGAACAGCACGTTTATCATTTAGAGAGACGCTCTCAG 1166
QY 1021 GTCTGAGAGACCCAGACCAAGTTTGCAGCCACTGTGGCCCTGCTGTGGCGGGAAGCTG 1080
DB 1167 GTCTGAGAGACCCAGACCAAGTTTGCAGCCACTGTGGCCCTGCTGTGGCGGGAAGCTG 1226
QY 1081 AAGCTGCACATGAACCCCGCCAGGTGAAGGCCACCATCATCATGAGCAGCAGGCCAAG 1140
DB 1227 AAGCTGCACATGAACCCCGCCAGGTGAAGGCCACCATCATCATGAGCAGCAGGCCAAG 1286
QY 1141 TCTCTGCTCAAGAACGAGAACACCCCGCAATGATTTACAGTGGCGAGATCTTTGAACAAC 1200
DB 1287 TCTCTGCTCAAGAACGAGAACACCCCGCAATGATTTACAGTGGCGAGATCTTTGAACAAC 1346
QY 1201 TGGCTCATGAGTACACCAAGCCACAGGCCACCCCTTAGTGGCCACATTCAGGAATATGTC 1260
DB 1347 TGGCTCATGAGTACACCAAGCCACAGGCCACCCCTTAGTGGCCACATTCAGGAATATGTC 1406
QY 1261 CTGAACCGAATTAAGAGGTGAGACCCGCTGTGGGCGCAGAGTCGGTGACAGAGAAAAATTT 1320
DB 1407 CTGAACCGAATTAAGAGGTGAGACCCGCTGTGGGCGCAGAGTCGGTGACAGAGAAAAATTT 1466
QY 1321 ACAAATCTGTTTGAATCCAGTTTCAAGTGTGGTGAAGTGAAGTGTGTTTCAAGTCAAG 1380
DB 1467 ACAAATCTGTTTGAATCCAGTTTCAAGTGTGGTGAAGTGAAGTGTGTTTCAAGTCAAG 1526
QY 1381 ACCCTGTCCTGCGCAGTGTGGTGTGATCGTTTCATGGCAGCCAGGACACATTCGACCGCC 1440
DB 1527 ACCCTGTCCTGCGCAGTGTGGTGTGATCGTTTCATGGCAGCCAGGACACATTCGACCGCC 1586
QY 1441 ACTGTTCTCTGGACCAATGCTTTTTCAGAGCTGCGCAGGCTGCGCATTTGCGCTGCTGAC 1500
DB 1587 ACTGTTCTCTGGACCAATGCTTTTTCAGAGCTGCGCAGGCTGCGCATTTGCGCTGCTGAC 1646
QY 1501 AAAGTGTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAAATTCAGAGCGGAGTGCAG 1560
DB 1647 AAAGTGTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAAATTCAGAGCGGAGTGCAG 1706
QY 1561 AGCAACCGGGGCTGACCAAGGAGAACCTCGTGTCTCTGGCGCAGAGAACCTGTTCAACAAC 1620
DB 1707 AGCAACCGGGGCTGACCAAGGAGAACCTCGTGTCTCTGGCGCAGAGAACCTGTTCAACAAC 1766
QY 1621 AGCAGCAGCCACCTGGAGGACTACAGTGGCTGTCTGTGCTCTGCTGCCAGTTCACACAGG 1680
DB 1767 AGCAGCAGCCACCTGGAGGACTACAGTGGCTGTCTGTGCTGCTGCTGCCAGTTCACACAGG 1826
QY 1681 GAGAAATTTACAGGACGGAATTTACATTTCTGGCAATGTTTGAACGTTGATGGAAGTG 1740
DB 1827 GAGAAATTTACAGGACGGAATTTACATTTCTGGCAATGTTTGAACGTTGATGGAAGTG 1886
QY 1741 TTAAGAAACATCTCAAGCCCTCATTTGAATGATGGGCGCCATTTTGGGCTTTGTTAAACAAG 1800

Db 1887 TTAATAAACATCTCAAGCCTCATTTGGAATGATGGGCCATTTTGGGTTTGTAAACAAG 1946
Qy 1801 CAACAGGCCCATGACCTACTATTAAACAAGCCAGATGGGACCTTCTCTGAGATTCAGT 1860
Db 1947 CAACAGGCCCATGACCTACTATTAAACAAGCCAGATGGGACCTTCTCTGAGATTCAGT 2006
Qy 1861 GACTCAGAAATTTGGCGGCATCACCATTTGCTTTGGAAATTTGATTTCTCAGAAAGAAATGTTT 1920
Db 2007 GACTCAGAAATTTGGCGGCATCACCATTTGCTTTGGAAATTTGATTTCTCAGAAAGAAATGTTT 2066
Qy 1921 TGGAAATCTGATGCTTTTACCAACAGAGACTTCTCCATCAGGTCCCTTAGCGACCGCTTG 1980
Db 2067 TGGAAATCTGATGCTTTTACCAACAGAGACTTCTCCATCAGGTCCCTTAGCGACCGCTTG 2126
Qy 1981 GGAGACTTGAATTAACCTTATCTACGTGTTTCTGATCGGCCCAAAAGATGAAGTATCTCC 2040
Db 2127 GGAGACTTGAATTAACCTTATCTACGTGTTTCTGATCGGCCCAAAAGATGAAGTATCTCC 2186
Qy 2041 AAATACTACACACAGATTTCCCTGCGAGTCTGCTACTGCTAAAGCTTTGATGGATACGTG 2100
Db 2187 AAATACTACACACAGATTTCCCTGCGAGTCTGCTACTGCTAAAGCTTTGATGGATACGTG 2246
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAAGATGCCGGGGC 2160
Db 2247 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAAGATGCCGGGGC 2306
Qy 2161 GGCAGCGCCAGTATACATGAGCAGGCCCCCTCCAGCTGTGTGTCGCCAGGCTCACTAT 2220
Db 2307 GGCAGCGCCAGTATACATGAGCAGGCCCCCTCCAGCTGTGTGTCGCCAGGCTCACTAT 2366
Qy 2221 AACATGTACCCACAGAACCTTGACTCAGTCTTGTGACACCGATGGGACTTCGATCTGGAG 2280
Db 2367 AACATGTACCCACAGAACCTTGACTCAGTCTTGTGACACCGATGGGACTTCGATCTGGAG 2426
Qy 2281 GACCAATATGAGCTAGCGCGCGGTGTGGAGGAGCTCTCTGGGCGCGCCCAATGGACAGTCAG 2340
Db 2427 GACCAATATGAGCTAGCGCGCGGTGTGGAGGAGCTCTCTGGGCGCGCCCAATGGACAGTCAG 2486
Qy 2341 TGGATCCCGCAGCGCAATCTGTA 2364
Db 2487 TGGATCCCGCAGCGCAATCTGTA 2510

RESULT 6
ADA02655
ID ADA02655 standard; cDNA; 2782 BP.
AC ADA02655;
XX
DT 06-NOV-2003 (first entry)
XX Human STAT5B carcinoma associated cDNA, SEQ ID NO:1173.
DE Human: carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.
XX Homo sapiens.
OS
XX WO2003057146-A2.
PN
XX 17-JUL-2003.
PD
XX 26-DEC-2002; 2002WO-US041414.
PF
XX 26-DEC-2001; 2001US-00035832.
PR
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Moirris DW;
XX WPI; 2003-587068/55.
DR

XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
FS Claim 1; SEQ ID NO 1173; 245pp; English.
XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 2782 BP; 669 A; 739 C; 801 G; 573 T; 0 U; 0 Other;
Query Match 100.0%; Score 2364; DB 9; Length 2782;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGCTGTGTGGATACAAAGCTCAGCAGCTCAAGAGAGAGAGCCCTTCATCAGATGCAAGCG 60
Db 147 ATGGCTGTGTGGATACAAAGCTCAGCAGCTCAAGAGAGAGAGCCCTTCATCAGATGCAAGCG 206
Qy 61 TTATATGCCAGCATTTTCCCATTTGAGTGGGCAATATTTATCCCATGGATTTGAAGC 120
Db 207 TTATATGCCAGCATTTTCCCATTTGAGTGGGCAATATTTATCCCATGGATTTGAAGC 266
Qy 121 CAAGCATGGGACTCAGTAGATCTTTGATAATCCAGAGAGAAACATTAAAGCCACCCAGCTC 180
Db 267 CAAGCATGGGACTCAGTAGATCTTTGATAATCCAGAGAGAAACATTAAAGCCACCCAGCTC 326
Qy 181 CTGGAGGGCTCTGTGTCAGGAGCTGCAGAAAGGAGCAGCAGCAGGTTGGGGAAGATGGG 240
Db 327 CTGGAGGGCTCTGTGTCAGGAGCTGCAGAAAGGAGCAGCAGCAGGTTGGGGAAGATGGG 386
Qy 241 TTTTACTGAAAGATCAAGCTGGGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC 300
Db 387 TTTTACTGAAAGATCAAGCTGGGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC 446
Qy 301 TGCCCCATGGAGCTGTCTCCGCTGATCGCCATATATTTGTAATGAAACAGAGTTTGTGTC 360
Db 447 TGCCCCATGGAGCTGTCTCCGCTGATCGCCATATATTTGTAATGAAACAGAGTTTGTGTC 506
Qy 361 CGAGAAGCCAAACAATGGTAGCTCTCCAGCTGGAAAGCCCTTGTGATGCCATGTGCCAGAAA 420
Db 507 CGAGAAGCCAAACAATGGTAGCTCTCCAGCTGGAAAGCCCTTGTGATGCCATGTGCCAGAAA 566
Qy 421 CACTCCAGATCAACCCAGACGTTTGGAGAGCTGGCA CTGGTCACTGCGAGGACACAGAGAT 480
Db 567 CACTCCAGATCAACCCAGACGTTTGGAGAGCTGGCA CTGGTCACTGCGAGGACACAGAGAT 626
Qy 481 GAGTTAAAGAGCTGCAGCAGACTCAGAGTACTTTCATCTCATCTCAGTACCAGAGAGCTG 540
Db 627 GAGTTAAAGAGCTGCAGCAGACTCAGAGTACTTTCATCTCATCTCAGTACCAGAGAGCTG 686
Qy 541 AGGATCCAAGCTCAGTTTGGGCCGCTGGCCCGAGCTGAGCCCCCAGGAGCGCTCTGAGCCGG 600

Db 687 AGGATCCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCCAGGAGCGTCTGAGCCGG 746
Qy 601 GAGACGGCCCTCCAGCAGAAAGAGGTGTCTCTGAGAGGCTGTGGTTGACGCTGAGGCAAG 660
Db 747 GAGACGGCCCTCCAGCAGAAAGAGGTGTCTCTGAGAGGCTGTGGTTGACGCTGAGGCAAG 806
Qy 661 ACATGTGACAGTAGTACCGGTGGAGCTGCCGAGAGCACCAGAAAGACCTCGACCTGCTG 720
Db 807 ACATGTGACAGTAGTACCGGTGGAGCTGCCGAGAGCACCAGAAAGACCTCGACCTGCTG 866
Qy 721 CGGAAGCAGCAGACCATCATCTCTGGATGACAGAGCTGATCCAGTGGAAAGCGCGCAGCAG 780
Db 867 CGGAAGCAGCAGACCATCATCTCTGGATGACAGAGCTGATCCAGTGGAAAGCGCGCAGCAG 926
Qy 781 CTGGCCGGGAAACGGCGGGCCCCCGAGGGGAGCCCTGAGAGCTGACGTGCTGAGTCTGCTG 840
Db 927 CTGGCCGGGAAACGGCGGGCCCCCGAGGGGAGCCCTGAGAGCTGCTGACAGTCTGCTGCTG 986
Qy 841 AAGTTGGCGGAGATCATCTGCGAGAAACGGCGAGCAGATCCGACGGCTGAGCACTCTGC 900
Db 987 AAGTTGGCGGAGATCATCTGCGAGAAACGGCGAGCAGATCCGACGGCTGAGCACTCTGC 1046
Qy 901 CAGCAGCTGCCATCCCGGGCCCCAGTGGAGGAGATGCTGGCGAGGTCAACGCCACCATC 960
Db 1047 CAGCAGCTGCCATCCCGGGCCCCAGTGGAGGAGATGCTGGCGAGGTCAACGCCACCATC 1106
Qy 961 ACGACATATTCTCAGCCCTGGTGACACGACAGTTTCATATTGAGAGAGCAGCTCTCTAG 1020
Db 1107 ACGACATATTCTCAGCCCTGGTGACACGACAGTTTCATATTGAGAGAGCAGCTCTCTAG 1166
Qy 1021 GTCTGTAAGACCCAGCTTTGACGCCACTGTGGCCCTGTGGTGGGGGGAAGCTG 1080
Db 1167 GTCTGTAAGACCCAGCTTTGACGCCACTGTGGCCCTGTGGTGGGGGGAAGCTG 1226
Qy 1081 AACGTGCATGAACCCCCCAGGTGAAGGCCACCATCATCAGTGAAGCAGCAGCCCAAG 1140
Db 1227 AACGTGCATGAACCCCCCAGGTGAAGGCCACCATCATCAGTGAAGCAGCAGCCCAAG 1286
Qy 1141 TCTCTGTCAAGAACGAGAAACACCGGCAATGATTAAGTGGCGAGATCTGAAACAATGC 1200
Db 1287 TCTCTGTCAAGAACGAGAAACACCGGCAATGATTAAGTGGCGAGATCTGAAACAATGC 1346
Qy 1201 TGGGTATGAGTAGTACCAAGCAGCAGCACCCTTATGTGCCACTTCAGAGAAATGTCC 1260
Db 1347 TGGGTATGAGTAGTACCAAGCAGCAGCACCCTTATGTGCCACTTCAGAGAAATGTCC 1406
Qy 1261 CTGAAAAGAAATTAAGAGGTACAGCCGTCTGGGGCAGAGTCGGTGACAGAGAAAAATTT 1320
Db 1407 CTGAAAAGAAATTAAGAGGTACAGCCGTCTGGGGCAGAGTCGGTGACAGAGAAAAATTT 1466
Qy 1321 ACAATCTCTGTTGAATCCAGTTTCAAGTTGGTGGAAATGAGCTGGTTTTCAGAGTCAAG 1380
Db 1467 ACAATCTCTGTTGAATCCAGTTTCAAGTTGGTGGAAATGAGCTGGTTTTCAGAGTCAAG 1526
Qy 1381 ACCCTGTCTCCAGTGGTGTGATCGTTCATGCGAGCAGGACCAATGCGACGCGC 1440
Db 1527 ACCCTGTCTCCAGTGGTGTGATCGTTCATGCGAGCAGGACCAATGCGACGCGC 1586
Qy 1441 ACTGTCTCTGGGCAATGCTTTTGCAGAGCCTGGCAGGGTGCCATTTCGCGTGCCTGAC 1500
Db 1587 ACTGTCTCTGGGCAATGCTTTTGCAGAGCCTGGCAGGGTGCCATTTCGCGTGCCTGAC 1646
Qy 1501 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGAGCCGAGAGTGAG 1560
Db 1647 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGAGCCGAGAGTGAG 1706
Qy 1561 AGCAACCGGGGCTTGACCAAGAGAACCTGTGTTCTTGGCGCAGAACTGTTCAACAAC 1620
Db 1707 AGCAACCGGGGCTTGACCAAGAGAACCTGTGTTCTTGGCGCAGAACTGTTCAACAAC 1766
Qy 1621 AGCAGACCCACCTGGAGGACTACAGTGGCCCTGTCTGTGCTCCAGTCCCAAGG 1680
Db 1767 AGCAGACCCACCTGGAGGACTACAGTGGCCCTGTCTGTGCTCCAGTCCCAAGG 1826

Qy 1581 GAGAAATTTACAGGACCGAATTTACACTTTCTGSCAATGTTTGGACGTGTGATGGAAGTG 1740
Db 1827 GAGAAATTTACAGGACCGAATTTACACTTTCTGSCAATGTTTGGACGTGTGATGGAAGTG 1886
Qy 1741 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTTTGTAAAACAAG 1800
Db 1887 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGGCCATTTTGGGGTTTGTAAAACAAG 1946
Qy 1801 CAACAGCCCATGACCTACTGATTTAAACAAGCCAGATGGGACCTTCTCTGAGATTCAAGT 1860
Db 1947 CAACAGCCCATGACCTACTGATTTAAACAAGCCAGATGGGACCTTCTCTGAGATTCAAGT 2006
Qy 1861 GACTCAGAAATTTGGCGGCATCCATTTGCTTTGGAGTTTGAATCTCAGGAAGAATGTTT 1920
Db 2007 GACTCAGAAATTTGGCGGCATCCATTTGCTTTGGAGTTTGAATCTCAGGAAGAATGTTT 2066
Qy 1921 TGGAAATCTGATGCTTTTACCAACAGAGATCTTCTCCATCAGGTCCCTAGCCGACCGTTG 1980
Db 2067 TGGAAATCTGATGCTTTTACCAACAGAGATCTTCTCCATCAGGTCCCTAGCCGACCGTTG 2126
Qy 1981 GGAGACTTTGAATTAACCTTATCTAGCTTTTCTGATCGGCCAAAAGATGAAGTATATCTCC 2040
Db 2127 GGAGACTTTGAATTAACCTTATCTAGCTTTTCTGATCGGCCAAAAGATGAAGTATATCTCC 2186
Qy 2041 AAATACTACACACAGTTCCTCGGAGTCTGCTACTTCTAAAGCTGTTGATGGATACGTG 2100
Db 2187 AAATACTACACACAGTTCCTCGGAGTCTGCTACTTCTAAAGCTGTTGATGGATACGTG 2246
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2160
Db 2247 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2306
Qy 2161 GGACAGCCCAAGTATACAGACAGGCCCTCCCTCCAGCTGTGTGTCCCAAGCTCACTAT 2220
Db 2307 GGACAGCCCAAGTATACAGACAGGCCCTCCCTCCAGCTGTGTGTCCCAAGCTCACTAT 2366
Qy 2221 AACATGTACCCACAGAACCCCTGACTCAGTCTCTTGCACCGATGGGACTTCGATCTGGAG 2280
Db 2367 AACATGTACCCACAGAACCCCTGACTCAGTCTCTTGCACCGATGGGACTTCGATCTGGAG 2426
Qy 2281 GACAAATGGACGCTAGCGCGCGTGTGGAGAGCTCTCTGGCGCGGCCAATGGAAGTCAAG 2340
Db 2427 GACAAATGGACGCTAGCGCGCGTGTGGAGAGCTCTCTGGCGCGGCCAATGGAAGTCAAG 2486
Qy 2341 TGGATCCCGCAGCAGCAATCGTGA 2364
Db 2487 TGGATCCCGCAGCAGCAATCGTGA 2510

RESULT 7

ADB72393

ID ADB72393 standard; mRNA; 2782 BP.

XX AC ADB72393;

XX AC ADB72393;

XX 04-DEC-2003 (first entry)

XX Human STAT5B mRNA.

XX human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;

XX cancer; neoplasm; adenocarcinoma; sarcoma.

XX Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

```
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW, Engelhard EK;
XX WPI; 2003-239337/23.
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX cancers, neoplasm, adenocarcinoma, or sarcomas.
XX Claim 1; SEQ ID NO 221; 2304pp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
XX nucleotide sequence selected from any of the 660 sequences fully defined
XX in the specification. A polynucleotide of the invention has cytostatic
XX activity, and may have a use in gene therapy, or in a vaccine. The
XX recombinant nucleic acids and polypeptides are useful for treating
XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX sarcomas. The present sequence represents a human mRNA of the invention.
XX
XX Sequence 2782 BP; 669 A; 739 C; 801 G; 573 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 2364; DB 10; Length 2782;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ATGGCTGTGGATCAAGCTCAGCAGCTCAAGAGAGAGCCCTTTCATCAGATCAAGCG 60
DB 147 ATGGCTGTGGATCAAGCTCAGCAGCTCAAGAGAGAGCCCTTTCATCAGATCAAGCG 206
QY 61 TTATATGCCAGCATTTTCCATTGAGTGGGCAATTTATCCAGTGGATTGAAGC 120
DB 207 TTATATGCCAGCATTTTCCATTGAGTGGGCAATTTATCCAGTGGATTGAAGC 266
QY 121 CAAGCATGGGACTCAGTAGACTTCTAATATCCAGAGAGACATTAAGGCCACCCAGCTC 180
DB 267 CAAGCATGGGACTCAGTAGACTTCTAATATCCAGAGAGACATTAAGGCCACCCAGCTC 326
QY 181 CTGAGGCGCTGGTCAGAGCTGCAGAGAGGAGAGAGCCAGCTGGGGAGAGATGGG 240
DB 327 CTGAGGCGCTGGTCAGAGCTGCAGAGAGGAGAGAGCCAGCTGGGGAGAGATGGG 386
QY 241 TTTTACTGAAGATCAAGCTGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC 300
DB 387 TTTTACTGAAGATCAAGCTGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC 446
QY 301 TGCCCATGGAGCTGGTCCGCTGCATCCGCATATATTGTAATGAACAGAGGTTGGTC 360
DB 447 TGCCCATGGAGCTGGTCCGCTGCATCCGCATATATTGTAATGAACAGAGGTTGGTC 506
QY 361 CGAAGAGCAACAATGGTAGCTCTCCAGCTGGAAGCCCTTGCTGATGCCATGTCCAGAAA 420
DB 507 CGAAGAGCAACAATGGTAGCTCTCCAGCTGGAAGCCCTTGCTGATGCCATGTCCAGAAA 566
QY 421 CACTTCCAGATCAACAGAGCTTTGAGGAGCTGCGACTGGTCAACGAGGACACAGAGAT 480
DB 567 CACTTCCAGATCAACAGAGCTTTGAGGAGCTGCGACTGGTCAACGAGGACACAGAGAT 626
QY 481 GAGTTAAAGAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACAGGAGGCTG 540
DB 627 GAGTTAAAGAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACAGGAGGCTG 686
QY 541 AGGATCCAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCGCCAGGAGGCTCTGAGCCGG 600
DB 687 AGGATCCAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCGCCAGGAGGCTCTGAGCCGG 746
QY 601 GAGAGCGCCCTCCAGCAGAGCAGGTGTCTCTGGAGGCTGGTTCAGCGGTGAGGACAG 660
DB 747 GAGAGCGCCCTCCAGCAGAGCAGGTGTCTCTGGAGGCTGGTTCAGCGGTGAGGACAG 806
```

Db 1887 TTAATAAATCTCAAGCCCTCATGGATGATGGGCCATTTTGGGTTTGTAAACAAG 1946
Qy 1801 CAACAGGCCCATGACCTACTAGTAATTAACAAGCCAGATGGGACCTTCTCTGAGATTCAGT 1860
Db 1947 CAACAGGCCCATGACCTACTAGTAATTAACAAGCCAGATGGGACCTTCTCTGAGATTCAGT 2006
Qy 1861 GACTCAGAAATTTGGGGCATCACCATTCCTTGGGAAGTTTGATTCCTCAGGAAGAAATCTTT 1920
Db 2007 GACTCAGAAATTTGGGGCATCACCATTCCTTGGGAAGTTTGATTCCTCAGGAAGAAATCTTT 2066
Qy 1921 TGGAAATCTGATGCCCTTTTACCACAGAGACTTCTCCATCAGGTCCTTACGCCGACCGCTTG 1980
Db 2067 TGGAAATCTGATGCCCTTTTACCACAGAGACTTCTCCATCAGGTCCTTACGCCGACCGCTTG 2126
Qy 1981 GGAGACTTGAATTAACCTTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATATCTCC 2040
Db 2127 GGAGACTTGAATTAACCTTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATATCTCC 2186
Qy 2041 AAATACTACACACAGTTCCCTGGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2100
Db 2187 AAATACTACACACAGTTCCCTGGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2246
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGATCTGCAGATGCCGGGGGC 2160
Db 2247 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGATCTGCAGATGCCGGGGGC 2306
Qy 2161 GGCAGGCCAGTACATGAGACACAGGCCCTTCCAGCTGTGTCTCCAGGCTCACTAT 2220
Db 2307 GGCAGGCCAGTACATGAGACACAGGCCCTTCCAGCTGTGTCTCCAGGCTCACTAT 2366
Qy 2221 AACATGTACCCACAGAACCTGACTCAGTCTCTGACACCGATGGGACTTGGATCTGGAG 2280
Db 2367 AACATGTACCCACAGAACCTGACTCAGTCTCTGACACCGATGGGACTTGGATCTGGAG 2426
Qy 2281 GACACAAATGGAGCTAGCGCGCGTGTGGAGGAGCTCTCTGGCGCGCAATGGACAGTCAG 2340
Db 2427 GACACAAATGGAGCTAGCGCGCGTGTGGAGGAGCTCTCTGGCGCGCAATGGACAGTCAG 2486
Qy 2341 TGGATCCCGCACGCGACAAATCTGTA 2364
Db 2487 TGGATCCCGCACGCGACAAATCTGTA 2510

RESULT 8

AD95903
ID ADE95903 standard; DNA; 2782 BP.

XX AC ADE95903;

XX DT 12-FEB-2004 (first entry)

XX DE Human DNA related to STAT5B gene mRNA.

XX KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
XX KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; STAT5B.

XX OS Homo sapiens.

XX FN WO2003039484-A2.

XX PD 15-MAY-2003.

XX PF 08-NOV-2002; 2002WO-US036071.

XX PR 08-NOV-2001; 2001US-00052482.

XX PA (SAGR-) SAGRES DISCOVERY.

XX FI Morris DW, Engelhard EK;

XX DR WPI; 2003-441462/41.

XX

PT New carcinoma associated nucleic acids and proteins, useful for screening
PT drug candidates, or for diagnosing and treating carcinomas, e.g.
PT lymphoma, breast cancer, prostate cancer or leukemia.

XX Claim 1; SEQ ID NO 161; 793pp; English.

XX This invention relates to novel recombinant nucleic acids for use in
CC diagnosis and treatment of cancer, especially carcinomas, as well as the
CC use of compositions in screening methods. The compositions of the
CC invention may have cytostatic activity whilst the disclosed sequences may
CC be useful for gene therapy. The carcinoma associated nucleic acids and
CC proteins are useful for diagnosing and treating carcinomas, for example
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC drug candidates or bioactive agents capable of binding to, or modulating
CC the activity of, a carcinoma associated protein. The present sequence is
CC a DNA sequence which represents the mRNA derived from the human STAT5B
CC gene which is a carcinoma associated gene of the invention.

XX Sequence 2782 BP; 669 A; 739 C; 801 G; 573 T; 0 U; 0 Other;

Query Match 100.0%; Score 2364; DB 10; Length 2782;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTGTGGATACAAAGCTCAGCAGCTCAAGGAGAAGCCCTTATCAGATGCAAGCG 60
Db 147 ATGGCTGTGGATACAAAGCTCAGCAGCTCAAGGAGAAGCCCTTATCAGATGCAAGCG 206
Qy 61 TTATATGGCCAGCATTTTCCATTTAGGTCGGCATTATTTATCCAGTGGATTGAAGC 120
Db 207 TTATATGGCCAGCATTTTCCATTTAGGTCGGCATTATTTATCCAGTGGATTGAAGC 266
Qy 121 CAAGCATGGAGCTCAGTAGATCTTGATTAATCCACAGAGAACATTAAGGCCACCCAGCTC 180
Db 267 CAAGCATGGAGCTCAGTAGATCTTGATTAATCCACAGAGAACATTAAGGCCACCCAGCTC 326
Qy 181 CTGAGGGCCCTGGTGCAGGAGCTGCAGAGAAGCAGACACAGGTGGGGAAGATGGG 240
Db 327 CTGAGGGCCCTGGTGCAGGAGCTGCAGAGAAGCAGACACAGGTGGGGAAGATGGG 386
Qy 241 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACA GTATGACCGC 300
Db 387 TTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGAAACA GTATGACCGC 446
Qy 301 TGCCCATGGAGCTGGTCCGCTGCATCCGCATATATTGTACATGACAGAGTTGGTC 360
Db 447 TGCCCATGGAGCTGGTCCGCTGCATCCGCATATATTGTACATGACAGAGTTGGTC 506
Qy 361 CGAGAAGCCAAACATGGTAGCTCTCCAGCTGGAGCCCTTGTGATGCCATGTCCCGAATA 420
Db 507 CGAGAAGCCAAACATGGTAGCTCTCCAGCTGGAGCCCTTGTGATGCCATGTCCCGAATA 566
Qy 421 CACCTCCAGATCAACAGACGTTTGGAGAGCTGCGACTGTGTCACGAGGACACAGAGAT 480
Db 567 CACCTCCAGATCAACAGACGTTTGGAGAGCTGCGACTGTGTCACGAGGACACAGAGAT 626
Qy 481 GAGTTAAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACCAGGAGCGCTG 540
Db 627 GAGTTAAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACCAGGAGCGCTG 686
Qy 541 AGGATCCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCGCCAGGAGCGCTCTCAGCGCG 600
Db 687 AGGATCCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCGCCAGGAGCGCTCTCAGCGCG 746
Qy 601 GAGACGGCCCTTCCAGCAGAAAGAGGTGTCTCTGAGGGCTTGGTTGTCAGCGTGAAGCAG 660
Db 747 GAGACGGCCCTTCCAGCAGAAAGAGGTGTCTCTGAGGGCTTGGTTGTCAGCGTGAAGCAG 806
Qy 661 ACATGACAGCTACCGCGTGGAGCTGCCCGAGAGCAACAGAGACCTTGCAGCTGCTG 720
Db 807 ACATGACAGCTACCGCGTGGAGCTGCCCGAGAGCAACAGAGACCTTGCAGCTGCTG 866
Qy 721 CGGAAGCAGCAGACCATCATCTCTGGATGACGAGCTGATCCAGTGGAAAGCGCGCGCAGCAG 780

Db 867 CGGAAGCAGCAGACCATCATCTGATGACGAGCTGATCCAGTGAAGCGCGGAGCAG 926
Qy 781 CTGGCCGGGAAACGGCGGCCCCCGAGGCGAGCTGCGTACAGTCTCTGGTGTGAG 840
Db 927 CTGGCCGGGAAACGGCGGCCCCCGAGGCGAGCTGCGTACAGTCTCTGGTGTGAG 986
Qy 841 AAGTTGGCGGAGATCATCTGGCAGAAACCGGCGAGAGATCCGCGAGGCTGAGCACCTCTGC 900
Db 987 AAGTTGGCGGAGATCATCTGGCAGAAACCGGCGAGAGATCCGCGAGGCTGAGCACCTCTGC 1046
Qy 901 CAGCAGCTGCCCATCCCCGGGCCAGTGGAGAGATGCTGGCGAGAGTCAACGCCCAACATC 960
Db 1047 CAGCAGCTGCCCATCCCCGGGCCAGTGGAGAGATGCTGGCGAGAGTCAACGCCCAACATC 1106
Qy 961 ACGGACATTAATCTCAGCCCTGGTGAACAGCAGCTTCAATGAGAACGAGCCTCCTCAG 1020
Db 1107 ACGGACATTAATCTCAGCCCTGGTGAACAGCAGCTTCAATGAGAACGAGCCTCCTCAG 1166
Qy 1021 GTCTGGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCTGCTGGTGGCGGGAAGCTG 1080
Db 1167 GTCTGGAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCTGCTGGTGGCGGGAAGCTG 1226
Qy 1081 AACGTGCAATGAACCCCCCGAGGTGAAGGCCAACATCATCAGTGAAGAGAGGCCCAAG 1140
Db 1227 AACGTGCAATGAACCCCCCGAGGTGAAGGCCAACATCATCAGTGAAGAGAGGCCCAAG 1286
Qy 1141 TCTCTGCTCAAGAACAGAACACCCCGCAATGATTAAGTGGGAGATCTTGAACAACTGC 1200
Db 1287 TCTCTGCTCAAGAACAGAACACCCCGCAATGATTAAGTGGGAGATCTTGAACAACTGC 1346
Qy 1201 TGCCTGAGGAGTACACCAAGCCACAGCACCCCTTAGTGGCCACTTCAGGAATATGTCC 1260
Db 1347 TGCCTGAGGAGTACACCAAGCCACAGCACCCCTTAGTGGCCACTTCAGGAATATGTCC 1406
Qy 1261 CTGAACCGAATTAAGAGGTCAAGCCGTCTGGGGCAGAGTGGTGAACAGAAAGAAATTT 1320
Db 1407 CTGAACCGAATTAAGAGGTCAAGCCGTCTGGGGCAGAGTGGTGAACAGAAAGAAATTT 1466
Qy 1321 ACAATCTCTTTGAATCCAGTTTCAAGTGTGGTGAATAGAGCTGGTTTTCAAGTCAAG 1380
Db 1467 ACAATCTCTTTGAATCCAGTTTCAAGTGTGGTGAATAGAGCTGGTTTTCAAGTCAAG 1526
Qy 1381 ACCCTGTCCCTGCCAGTGTGTGATCTTCAATGCGAGCAGGACCAATGCGAGGCC 1440
Db 1527 ACCCTGTCCCTGCCAGTGTGTGATCTTCAATGCGAGCAGGACCAATGCGAGGCC 1586
Qy 1441 ACTGTTCTCTGGGACAAATGCTTTTGCAGAGCCTGGCAGGGTGCCATTTGCCGTGCTGAC 1500
Db 1587 ACTGTTCTCTGGGACAAATGCTTTTGCAGAGCCTGGCAGGGTGCCATTTGCCGTGCTGAC 1646
Qy 1501 AAGTGTCTGGGCCACAGCTGTGAGGGCTCAACATGAATTCAGAGCGGAGAGTGAG 1560
Db 1647 AAGTGTCTGGGCCACAGCTGTGAGGGCTCAACATGAATTCAGAGCGGAGAGTGAG 1706
Qy 1561 AGCAACGGGCGCTGACCAAGAGAGAACCTGGTTCTGGCGCAGAACTGTTCAACAAAC 1620
Db 1707 AGCAACGGGCGCTGACCAAGAGAGAACCTGGTTCTGGCGCAGAACTGTTCAACAAAC 1766
Qy 1621 AGCAGCAGCCACTGGAGACTTACAGTGGCCCTGTCTGTCTGGTCCAGTTTCAACAGG 1680
Db 1767 AGCAGCAGCCACTGGAGACTTACAGTGGCCCTGTCTGTCTGGTCCAGTTTCAACAGG 1826
Qy 1681 GAGAAATTTACAGGACGGAATTAACATTTCTGGCAATGTTTGAACGGTGTGAAGTGTG 1740
Db 1827 GAGAAATTTACAGGACGGAATTTACATTTCTGGCAATGTTTGAACGGTGTGAAGTGTG 1886
Qy 1741 TTAATAAACATCTCAGCCTCATTTGGAATGATGGGGCCATTTGGGGTTTGTAAACAAAG 1800
Db 1887 TTAATAAACATCTCAGCCTCATTTGGAATGATGGGGCCATTTGGGGTTTGTAAACAAAG 1946
Qy 1801 CAACAGGCCCATGACTTGAATTAACAGCCAGATGGACCTTCTCTGAGATTCAGT 1860

Db 1947 CAACAGGCCCATGACTTACTGATTAACAAAGCCAGATGGGACCTTCTCTGAGATTCAGT 2006
Qy 1861 GACTCAGAAAATTGGCGGCATCACCATTTGCTTTGGAAATTTGATTTCTCAGAAAGAAATGTTT 1920
Db 2007 GACTCAGAAAATTGGCGGCATCACCATTTGCTTTGGAAATTTGATTTCTCAGAAAGAAATGTTT 2066
Qy 1921 TGGAAATCTGATGCTCTTTTACACAGAGACTTCTCCATCAGGTCCCTAGCCGCTTG 1980
Db 2067 TGGAAATCTGATGCTCTTTTACACAGAGACTTCTCCATCAGGTCCCTAGCCGCTTG 2126
Qy 1981 GGAGACTTGAATTAACCTTTATCTAGTGTCTTCTGATCGGCCAAAAGATGAAGTATCTCC 2040
Db 2127 GGAGACTTGAATTAACCTTTATCTAGTGTCTTCTGATCGGCCAAAAGATGAAGTATCTCC 2186
Qy 2041 AAATACTACACACAGATTTCCCTGCGAGTCTCTACTCTGCTAAAGCTTTGATGGATACGTG 2100
Db 2187 AAATACTACACACAGATTTCCCTGCGAGTCTCTACTCTGCTAAAGCTTTGATGGATACGTG 2246
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCCCGGGGC 2160
Db 2247 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCCCGGGGC 2306
Qy 2161 GGCAGCGCCACGTATACAGGACAGCCCTCCCTCCAGCTGTGTGCCAGGCTCACTAT 2220
Db 2307 GGCAGCGCCACGTATACAGGACAGCCCTCCCTCCAGCTGTGTGCCAGGCTCACTAT 2366
Qy 2221 AACATGTACCCACAGAACCTGACTCAGTCTCTTGACACCGAGTGGGACTTCGATCTGGAG 2280
Db 2367 AACATGTACCCACAGAACCTGACTCAGTCTCTTGACACCGAGTGGGACTTCGATCTGGAG 2426
Qy 2281 GACCAATGGACGTAGCGCGGCTGTGGAGGAGCTCTCTGGGCGGCCCAATGGACAGTCAAG 2340
Db 2427 GACCAATGGACGTAGCGCGGCTGTGGAGGAGCTCTCTGGGCGGCCCAATGGACAGTCAAG 2486
Qy 2341 TGGATCCCGCACGCAACATCTGTA 2364
Db 2487 TGGATCCCGCACGCAACATCTGTA 2510

RESULT 9
ADP82726
ID ADP82726 standard; DNA; 2364 BP.
XX AC ADP82726;
XX DT 12-AUG-2004 (first entry)
XX Human STAT5B protein encoding sequence.
DE pluriptency; signal transducer and activator of transcription 5; STAT5;
KW Cytostatic; Dermatological; Antiinflammatory; Antithyroid; Antidiabetic;
KW Hepatotropic; Nephrotropic; Cardiant; prophylaxis; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 1..2364
XX FT /*tag= a
XX FT /product= "STAT5B"
XX PN WO2004042040-A1.
XX XX 21-MAY-2004.
XX XX 23-OCT-2003; 2003WO-JP013579.
XX XX 08-NOV-2002; 2002JP-00326190.
XX XX (REPR-) REPROCELL INC.
XX PA Nakauchi H, Iwama A;
XX PI WPI; 2004-411515/38.
XX DR

DR P-PSDB; ADP82727.
XX
PT New composition comprising active STAT5 or an agent capable of activating
PT STAT5, useful for maintaining the expansion, pluripotency, or self-
XX replication ability of a stem cell.
XX
PS Claim 5; SEQ ID NO 3; 249pp; English.
XX
CC The present invention relates to a composition for maintaining the
CC expansion, pluripotency, or self-replication ability of a stem cell
CC comprises active signal transducer and activator of transcription 5
CC (STAT5), an agent capable of activating STAT5, and a nucleic acid
CC molecule encoding active STAT5. The STAT5, active STAT5, or an agent
CC capable of activating STAT5 is useful for maintaining the expansion,
CC pluripotency, or self-replication ability of a stem cell. The cell
CC obtained by treating a stem cell with active STAT5, or STAT5 and an agent
CC capable of activating the STAT5 is useful for the treatment or
CC prophylaxis of a disease or disorder in need of a stem cell or a
CC differentiated cell derived from it, or for preparing a medicament for
CC the treatment or prophylaxis of the disease or disorder. Diseases include
CC atrichia, melanoma, cutis malignant lymphoma, hemangiosarcoma,
CC histiocytosis, hydroa, pustulosis, dermatitis, or eczema. It also
CC includes hypothalamus/hypophysis diseases, thyroid gland diseases,
CC accessory thyroid gland diseases, lipid metabolism abnormality, inborn
CC error of metabolism, hyperbilirubinemia, hyperbilirubinuria, mast cell
CC deficiency, diabetes insipidus, dwarfism, or Wolman's disease. It further
CC includes liver cirrhosis, pancreatitis, pancreas cancer, kidney diseases,
CC bladder diseases, male/female genital organ disease, heart failure, or
CC myocardial infarction. The present sequence represents STAT5B protein
CC encoding sequence.
XX
SQ Sequence 2364 BP; 593 A; 637 C; 565 G; 469 T; 0 U; 0 Other;

Query Match 99.7%; Score 2356; DB 12; Length 2364;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2359; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGCGTGTGTGATACAGCTCAGAGCTCAGAGCTCAAGAGAGAGCCCTTCATCAGATGCAAGCG 60
DB 1 ATGCGTGTGTGTGATACAGCTCAGAGCTCAGAGCTCAAGAGAGAGCCCTTCATCAGATGCAAGCG 60
QY 61 TTATATGCCAGCATTTTCCCATTTAGGTGGGATTTATTTATCCAGTGATTTGAAGC 120
DB 61 TTATATGCCAGCATTTTCCCATTTAGGTGGGATTTATTTATCCAGTGATTTGAAGC 120
QY 121 CAAGCATGGGACTCAGTAGATCTTTGATTAATCCACAGGAGAACTTAAGGCCACCCAGCTC 180
DB 121 CAAGCATGGGACTCAGTAGATCTTTGATTAATCCACAGGAGAACTTAAGGCCACCCAGCTC 180
QY 181 CTGAGGGCCCTGGTGCAGGAGCTCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
DB 181 CTGAGGGCCCTGGTGCAGGAGCTCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
QY 241 TTTTCTACGAGATCAAGCTCGGACATATCCACAGCTCCAGAACACCTATGACCGC 300
DB 241 TTTTCTACGAGATCAAGCTCGGACATATCCACAGCTCCAGAACACCTATGACCGC 300
QY 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCGATATTTGTACAAATGAACAGAGGTGGTGC 360
DB 301 TGCCCCATGGAGCTGGTCCGCTGCATCCGCGATATTTGTACAAATGAACAGAGGTGGTGC 360
QY 361 CGAGAGCCAAATAGTGTAGCTCTCCAGCTGGAGAGCTTGTGTATGATGATGATGATGATGATG 420
DB 361 CGAGAGCCAAATAGTGTAGCTCTCCAGCTGGAGAGCTTGTGTATGATGATGATGATGATGATG 420
QY 421 CACCTCCAGATCAACAGACGTTTGAAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG 480
DB 421 CACCTCCAGATCAACAGACGTTTGAAGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG 480
QY 481 GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACCAGGAGAGCGCTG 540
DB 481 GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATCATCCAGTACCAGGAGAGCGCTG 540

QY 541 AGATCCAAAGCTCAGTTTGGCCCGCTGGCCCGAGCTGAGCCCCCAGGAGCGCTCTGAGCGG 600
DB 541 AGATCCAAAGCTCAGTTTGGCCCGCTGGCCCGAGCTGAGCCCCCAGGAGCGCTCTGAGCGG 600
QY 601 GAGACGGCCCTCCAGCAGAAAGCAGGTGTCTCTGGAGCGCTGGTTGCAAGCGTGAGGCAAG 660
DB 601 GAGACGGCCCTCCAGCAGAAAGCAGGTGTCTCTGGAGCGCTGGTTGCAAGCGTGAGGCAAG 660
QY 661 ACATGACAGCTACCCGCTGGAGCTGCCCGAGAGCAGCAGAGAGCCCTGACAGCTGCTG 720
DB 661 ACATGACAGCTACCCGCTGGAGCTGCCCGAGAGCAGCAGAGAGCCCTGACAGCTGCTG 720
QY 721 CGGAAGCAGCAGACCATCATCTCTGGATGACAGCTGATCAGTGGAAAGCGCGCAGCAG 780
DB 721 CGGAAGCAGCAGACCATCATCTCTGGATGACAGCTGATCAGTGGAAAGCGCGCAGCAG 780
QY 781 CTGSCCGGGAAACGGCGGGCCCCCGAGGGCAGCTTGAACGCTGCTACAGTCTCTGCTGAG 840
DB 781 CTGSCCGGGAAACGGCGGGCCCCCGAGGGCAGCTTGAACGCTGCTACAGTCTCTGCTGAG 840
QY 841 AAGTTGGCGGAGATCATCTGGCAGAACCGGACAGCAGATCCGAGGGCTGACACCTCTGC 900
DB 841 AAGTTGGCGGAGATCATCTGGCAGAACCGGACAGCAGATCCGAGGGCTGACACCTCTGC 900
QY 901 CAGCAGCTGCCATCCCGGCCCGAGGAGATGCTGGCCGAGGTCAACGCCACCATC 960
DB 901 CAGCAGCTGCCATCCCGGCCCGAGGAGATGCTGGCCGAGGTCAACGCCACCATC 960
QY 961 ACGGACATATTCTCAGCCCTGGTGACACGACGCTTCATCATTTGAGAGCAGCTCTCTCAG 1020
DB 961 ACGGACATATTCTCAGCCCTGGTGACACGACGCTTCATCATTTGAGAGCAGCTCTCTCAG 1020
QY 1021 GTCTGAAAGCCAGACCAAGTTTGGCAGCACTGTGCGCTGCTGGTGGGGGGGAAAGCTG 1080
DB 1021 GTCTGAAAGCCAGACCAAGTTTGGCAGCACTGTGCGCTGCTGGTGGGGGGGAAAGCTG 1080
QY 1081 AAGCTGACATGAACCCCGCCAGGTGAAGGCCACCATCATCATGTGAGCAGCAGGCGCAAG 1140
DB 1081 AAGCTGACATGAACCCCGCCAGGTGAAGGCCACCATCATCATGTGAGCAGCAGGCGCGCAAG 1140
QY 1141 TCTCTGCTCAAGAACAGAGAACCCCGCAATGATTTACAGTGGCGAGATCTTTGAACAACATGC 1200
DB 1141 TCTCTGCTCAAGAACAGAGAACCCCGCAATGATTTACAGTGGCGAGATCTTTGAACAACATGC 1200
QY 1201 TGCGTATGAGATACCAAGCCACAGGACCCCTTAGTGCCCTCCTCAGGAATATGTCC 1260
DB 1201 TGCGTATGAGATACCAAGCCACAGGACCCCTTAGTGCCCTCCTCAGGAATATGTCC 1260
QY 1261 CTGAAACGAAATTAGAGGCTCAGACCCGCTGTGGGCGAGGTCGTCAGAGAGAGAAATTT 1320
DB 1261 CTGAAACGAAATTAGAGGCTCAGACCCGCTGTGGGCGAGGTCGTCGTCAGAGAGAGAAATTT 1320
QY 1321 ACAATCTCTGTTGAATCCCAAGTTTCAAGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1380
DB 1321 ACAATCTCTGTTGAATCCCAAGTTTCAAGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1380
QY 1381 ACCGTGCTCTGCGCAGTGGTGGTATCTTCATGCGAGCGAGGACCAATGCGACGCGC 1440
DB 1381 ACCGTGCTCTGCGCAGTGGTGGTATCTTCATGCGAGCGAGGACCAATGCGACGCGC 1440
QY 1441 ACTGTTCTCTGGGCAATGCTTTTGCAGAGCTGGCAGGGTGCCATTTGCGCTGCTGAC 1500
DB 1441 ACTGTTCTCTGGGCAATGCTTTTGCAGAGCTGGCAGGGTGCCATTTGCGCTGCTGAC 1500
QY 1501 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTTCAAGGCCGAAAGTGCAG 1560
DB 1501 AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTTCAAGGCCGAAAGTGCAG 1560
QY 1561 AGCAACCGGGGCTGACCAAGGAGAACTGTGTCTTCTGGCGGAGAACTGTTCAACAAAC 1620
DB 1561 AGCAACCGGGGCTGACCAAGGAGAACTGTGTCTTCTGGCGGAGAACTGTTCAACAAAC 1620
QY 1621 AGCAGCAGCCACCTGGAGGACTACAGTGGGCTGCTGTGCTCTCCAGTTCACAGG 1680

Db 1621 AGCAGCACCACCTGGAGGACTACAGTGGCTGTCTGTCTCCAGTTCACACAGG 1680
Qy 1681 GAGAAATTTACAGACGGAATTTACACTTTCTGGCAATGGTTTGACGGTGTGATGGAAGTG 1740
Db 1691 GAGAAATTTACAGACGGAATTTACACTTTCTGGCAATGGTTTGACGGTGTGATGGAAGTG 1740
Qy 1741 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGCCAATTTGGGGTTTGTAAACAAG 1800
Db 1741 TTAATAAAACATCTCAAGCCTCATTTGGAATGATGGGCCAATTTGGGGTTTGTAAACAAG 1800
Qy 1801 CAACAGGCCCATGACCTACTGATTAACAGCCAGATGGGACCTTCTCTCTGAGATTTCAGT 1860
Db 1801 CAACAGGCCCATGACCTACTGATTAACAGCCAGATGGGACCTTCTCTCTGAGATTTCAGT 1860
Qy 1861 GACTCAGAAATTTGGGGCATCACCATTCTCTGGAGTTTGTATCTCAGGAAGATGTTT 1920
Db 1861 GACTCAGAAATTTGGGGCATCACCATTCTCTGGAGTTTGTATCTCAGGAAGATGTTT 1920
Qy 1921 TGGAAATCTGATGCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCCGACGGCTTG 1980
Db 1921 TGGAAATCTGATGCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCCGACGGCTTG 1980
Qy 1981 GGAGACTTGAATTAACCTTATCTAGTGTTCCTGATGGCCAAAGATGAAGTATATCTCC 2040
Db 1981 GGAGACTTGAATTAACCTTATCTAGTGTTCCTGATGGCCAAAGATGAAGTATATCTCC 2040
Qy 2041 AAATACTACACACAGTTCCCTGGAGTCTGCTACTGTAAGCTGTTGATGGATACGTG 2100
Db 2041 AAATACTACACACAGTTCCCTGGAGTCTGCTACTGTAAGCTGTTGATGGATACGTG 2100
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2160
Db 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2160
Qy 2161 GGCAGCGCCAGTACATGACAGCCGCCCTCCAGCTGTGTGTCCCGAGCTCACTAT 2220
Db 2161 GGCAGCGCCAGTACATGACAGCCGCCCTCCAGCTGTGTGTCCCGAGCTCACTAT 2220
Qy 2221 AACATGTACCCACAGAACCTTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Db 2221 AACATGTACCCACAGAACCTTGACTCAGTCTTGACACCGATGGGACTTCGATCTGGAG 2280
Qy 2281 GACAAATGGAGTAGCGCGCTGTGGAGAGCTCTCTGGCCGCGCAATGCAGATCAG 2340
Db 2281 GACAAATGGAGTAGCGCGCTGTGGAGAGCTCTCTGGCCGCGCAATGCAGATCAG 2340
Qy 2341 TGGATCCCGCAGCACAATCTGTA 2364
Db 2341 TGGATCCCGCAGCACAATCTGTA 2364

RESULT 10
ABZ35008
ID ABZ35008 standard; cDNA; 2716 BP.

XX AC ABZ35008;

XX DT 05-FEB-2003 (first entry)

XX DE Human gene expression profile polynucleotide SEQ ID NO 120.

XX KW Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
XX KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
XX KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
XX KW gene expression; gene; ss.

XX OS Homo sapiens.

XX FN WO200274979-A2.

XX PD 26-SEP-2002.

XX

PF 20-MAR-2002; 2002WO-US008456.
PR 20-MAR-2001; 2001US-0276947P.
PA (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
XX Wan J, Wang Y;
PI WPI; 2002-740862/80.
DR
XX
XX
PT New gene expression profile generated from primary, endothelial,
PT epithelial, and muscle cell types, useful for identifying disease
PT pathologies involving alterations of gene expression, e.g. cancer.
PS Claim 3; Page 305-306; 850pp; English.
XX

CC The invention relates to a gene expression profile comprising one or more
CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type
CC is a coronary artery endothelium, umbilical artery or vein endothelium,
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery
CC endothelium, myometrium microvascular endothelium, keratinocyte
CC epithelium, bronchial epithelium, mammary epithelium, prostate
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,
CC small airway epithelium, renal epithelium, umbilical artery smooth
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
CC osteoblasts or prostate stromal cell. The gene expression profile is used
CC for determining the level of RNA expression for a sample, determining the
CC phenotype of a cell and distinguishing cell types. The gene or a protein
CC expression profile is useful in identifying disease pathologies involving
CC alterations of gene expression. The assessment of expression profiles may
CC provide meaningful information with respect to tumour type and stage,
CC treatment methods, and prognosis. The gene or protein expression profile
CC may also be used for creating microarrays. The microarray is useful for
CC genetic and physical mapping of genomes, DNA sequencing, genetic or
CC medical diagnosis, genotyping of organisms, confirming cell or tissue
CC identifications and in identifying promising antibiotics, antiviral or
CC antifungal agents
XX

SQ Sequence 2716 BP; 662 A; 707 C; 760 G; 587 T; 0 U; 0 Other;

Query Match 99.7%; Score 2356; DB 6; Length 2716;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2359; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATGGCTGTGTGGATACAAAGCTCAGCAGCTCCAGGAGAAGCCCTTCATCAGATGCAAGCG 60
Db 12 ATGGCTGTGTGGATACAAAGCTCAGCAGCTCCAGGAGAAGCCCTTCATCAGATGCAAGCG 71
Qy 61 TTATATGGCCAGCATTTTCCATTGAGTGGGGCATTTATTTATCCAGTGGATTGAAAGC 120
Db 72 TTATATGGCCAGCATTTTCCATTGAGTGGGGCATTTATTTATCCAGTGGATTGAAAGC 131
Qy 121 CAGCAGTGGGACTCAGTAGATCTTGATAATCCAGGAGAACATTAAAGCCACCAGCTC 180
Db 132 CAAGCATGGGACTCAGTAGATCTTGATAATCCAGGAGAACATTAAAGCCACCAGCTC 191
Qy 181 CTGAGGGCCCTGGTGCAGGAGCTGCAGAAGAGCAGACACAGGTGGGGAAGATGGG 240
Db 192 CTGAGGGCCCTGGTGCAGGAGCTGCAGAAGAGCAGACACAGGTGGGGAAGATGGG 251
Qy 241 TTTTACTGAAGATCAAGCTGGGGCATTATGCCACACAGCTCCAGAACACATATGACCGC 300
Db 252 TTTTACTGAAGATCAAGCTGGGGCATTATGCCACACAGCTCCAGAACACATATGACCGC 311
Qy 301 TGCCCCATGGAGCTGGTCCCGTCGATCCGCATATATTGTACATGAACAGAGTTGGTC 360
Db 312 TGCCCCATGGAGCTGGTCCCGTCGATCCGCATATATTGTACATGAACAGAGTTGGTC 371
Qy 361 CGAGAAGCCCAACAAATGGTAGTCTCTCCAGCTGGGAAGCCCTTGTGTATGTCATGTCCAGAAA 420

Db 372 CGAAGCAACAAATGCTAGCTCTCCAGCTGGAAGCCTTGCTGATGCCATGTCTCCAGAAA 431
Qy 421 CACCTCCAGATCAACACAGACGTTTGGAGAGCTGCGACTGGTCAACGACGACACAGAGAAAT 480
Db 432 CACCTCCAGATCAACACAGACGTTTGGAGAGCTGCGACTGGTCAACGACGACACAGAGAAAT 491
Qy 481 GAGTTAAAAAGCTGACGACGACTCAGAGATPACTTTCATCATCTCCATGATACCAAGGAGAGCCTG 540
Db 492 GAGTTAAAAAGCTGACGACGACTCAGAGATPACTTTCATCATCTCCATGATACCAAGGAGAGCCTG 551
Qy 541 AGGATCAAGCTCAGTTTGGCCCGCTGGCCCGAGCTGAGCCCCCAGAGAGCGCTGAGAGCCGG 600
Db 552 AGGATCAAGCTCAGTTTGGCCCGCTGGCCCGAGCTGAGCCCCCAGAGAGCGCTGAGAGCCGG 611
Qy 601 GAGACGCGCTCCAGCAGAGACAGGTGTCTCTGAGAGCGCTGGTGGACGCTGAGGACAG 660
Db 612 GAGACGCGCTCCAGCAGAGACAGGTGTCTCTGAGAGCGCTGGTGGACGCTGAGGACAG 671
Qy 661 ACCTGACGAGTACCGCGTGGAGCTGCCGAGAGACACAGAGAGCCCTGCACTGCTG 720
Db 672 ACCTGACGAGTACCGCGTGGAGCTGCCGAGAGACACAGAGAGCCCTGCACTGCTG 731
Qy 721 CGGAAGCAGCAGACCATCATCTCTGGATGACAGAGCTGATCCAGTGGGAAGCGCGGCGACAG 780
Db 732 CGGAAGCAGCAGACCATCATCTCTGGATGACAGAGCTGATCCAGTGGGAAGCGCGGCGACAG 791
Qy 781 CTGCGCGGGAAACGCGCGGGCCCCCGAGGAGACCTGCAAGAGACCTGCACTGCTGAG 840
Db 792 CTGCGCGGGAAACGCGCGGGCCCCCGAGGAGACCTGCAAGAGACCTGCACTGCTGAG 851
Qy 841 AAGTTGCGGAGATCATCTGSCAGAACCGGCAGCAGATCCGAGGGCTGAGCACTCTGC 900
Db 852 AAGTTGCGGAGATCATCTGSCAGAACCGGCAGCAGATCCGAGGGCTGAGCACTCTGC 911
Qy 901 CAGCAGCTGCCATFCCCGGCCAGTGGAGAGATGCTGGCCGAGGTCAACGCCACCATC 960
Db 912 CAGCAGCTGCCATFCCCGGCCAGTGGAGAGATGCTGGCCGAGGTCAACGCCACCATC 971
Qy 961 ACGGACATTAATCTCAGCCCTGGTGACCAACGACGTTTCACTATGAGAGACGACCTCTCAG 1020
Db 972 ACGGACATTAATCTCAGCCCTGGTGACCAACGACGTTTCACTATGAGAGACGACCTCTCAG 1031
Qy 1021 GTCTGTAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCTGCTGGTGGCGGGAAGCTG 1080
Db 1032 GTCTGTAAGACCCAGACCAAGTTTGCAGCCACTGTGCGCTGCTGGTGGCGGGAAGCTG 1091
Qy 1081 AACGTGCAATGAACCCCCCCCCAGGTGAAGGCCACCATCATCAGTGAAGCAGAGGCCAAG 1140
Db 1092 AACGTGCAATGAACCCCCCCCCAGGTGAAGGCCACCATCATCAGTGAAGCAGAGGCCAAG 1151
Qy 1141 TCTCTGCTCAAGAACGAGAACACCCGCAATGATTAAGTGGCGAGATCTTGAACAAGTGC 1200
Db 1152 TCTCTGCTCAAGAACGAGAACACCCGCAATGATTAAGTGGCGAGATCTTGAACAAGTGC 1211
Qy 1201 TGGCTCATGGAGTACCAACCAAGCCACAGGACCTTAGTGGCCCACTTCAGGAATATGTCC 1260
Db 1212 TGGCTCATGGAGTACCAACCAAGCCACAGGACCTTAGTGGCCCACTTCAGGAATATGTCC 1271
Qy 1261 CTGAACCGAAATTAAGAGGTGAGACCGCTCGTGGGCGAGAGCTGGTGAAGAGAAAAATTT 1320
Db 1272 CTGAACCGAAATTAAGAGGTGAGACCGCTCGTGGGCGAGAGCTGGTGAAGAGAAAAATTT 1331
Qy 1321 ACAATCTGTTTGAATCCAGTTTCAAGTTGAGTGGGAAATGAGCTGGTTTTCAGAGTCAAG 1380
Db 1332 ACAATCTGTTTGAATCCAGTTTCAAGTTGAGTGGGAAATGAGCTGGTTTTCAGAGTCAAG 1391
Qy 1381 ACCCTGTCCCTGCGCAGTGGTGTATCGTTTCATGGCAGCCAGGACCAATGCGACGGCC 1440
Db 1392 ACCCTGTCCCTGCGCAGTGGTGTATCGTTTCATGGCAGCCAGGACCAATGCGACGGCC 1451
Qy 1441 ACTGTTCTCTGGGCAATAGCTTTTGCAGAGCTGGCAGGGTGCCATTTGCGCTGCTGAC 1500
Db 1452 ACTGTTCTCTGGGCAATAGCTTTTGCAGAGCTGGCAGGGTGCCATTTGCGCTGCTGAC 1511

Qy 1501 AAAGTCTGTGGCCACAGCTGTCTGAGGCGCTCAACATGAAATTCAGGCGCGAGTGCAG 1560
Db 1512 AAAGTCTGTGGCCACAGCTGTCTGAGGCGCTCAACATGAAATTCAGGCGCGAGTGCAG 1571
Qy 1561 AGCAACCGGGCGCTGACCAAGGAGAACCTCGTGTCTCTGGCGCAGAGAACTGTTCACAAC 1620
Db 1572 AGCAACCGGGCGCTGACCAAGGAGAACCTCGTGTCTCTGGCGCAGAGAACTGTTCACAAC 1631
Qy 1621 AGCAGCAGCCACTCTGAGGACTACAGTGGCTGTCTGTCTCTGGTCCCAAGTTCAACAG 1680
Db 1632 AGCAGCAGCCACTCTGAGGACTACAGTGGCTGTCTGTCTCTGGTCCCAAGTTCAACAG 1691
Qy 1681 GAGAAATTTACAGGAGCGAATTAACATTTCTGCGCAATGGTTTTCAGCGGTGTGATGGAAGTG 1740
Db 1692 GAGAAATTTACAGGAGCGAATTAACATTTCTGCGCAATGGTTTTCAGCGGTGTGATGGAAGTG 1751
Qy 1741 TTAATAAAAAATCTCAAGCCTCATTTGGAATGATGGGCCCAATTTTGGGGTTTGTAAACAAG 1800
Db 1752 TTAATAAAAAATCTCAAGCCTCATTTGGAATGATGGGCCCAATTTTGGGGTTTGTAAACAAG 1811
Qy 1801 CAACAGGCCCATGACCTTACCTGATTAACAAGCCAGATGGGACCTTCTCTCTGAGATTCACT 1860
Db 1812 CAACAGGCCCATGACCTTACCTGATTAACAAGCCAGATGGGACCTTCTCTCTGAGATTCACT 1871
Qy 1861 GACTCAGAAATTCGGCGCATCACCATTTGCTTGGAAATTTTGAATTTCTCAGGAAAGAAATGTTT 1920
Db 1872 GACTCAGAAATTCGGCGCATCACCATTTGCTTGGAAATTTTGAATTTCTCAGGAAAGAAATGTTT 1931
Qy 1921 TGGAAATCTGATGCTTTTACCACACAGAGACTTCTCCATCAGGTCCCTAGCCGACCGCTTG 1980
Db 1932 TGGAAATCTGATGCTTTTACCACACAGAGACTTCTCCATCCGGTCCCTAGCCGACCGCTTG 1991
Qy 1981 GGAAGCTTGAATTAACCTTATCTACGTGTTTCTGATCGGCCAAAAGATGAAGTATATCTCC 2040
Db 1992 GGAAGCTTGAATTAACCTTATCTACGTGTTTCTGATCGGCCAAAAGATGAAGTATATCTCC 2051
Qy 2041 AATATCTACACACAGTTCCTCGGAGTCTGCTACTGCTAAAGCTGTTGATGATACGTG 2100
Db 2052 AATATCTACACACAGTTCCTCGGAGTCTGCTACTGCTAAAGCTGTTGATGATACGTG 2111
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2160
Db 2112 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGC 2171
Qy 2161 GGCAGGCGCACGTATGAGACAGGCCCTCCCGACGTGTGTGTCCCGAGGCTCACTAT 2220
Db 2172 GGCAGGCGCACGTATGAGACAGGCCCTCCCGACGTGTGTGTCCCGAGGCTCACTAT 2231
Qy 2221 AACATGTACCCACAGAACCTGACTCAGTCTTTCAGACCGATGGGACTTTCGATCTGGAG 2280
Db 2232 AACATGTACCCACAGAACCTGACTCAGTCTTTCAGACCGATGGGACTTTCGATCTGGAG 2291
Qy 2281 GACACAATGGACGTAGCGCGCGTGTGGAGAGCTCTCTGGGCGGCGCAATGGACAGTCAAG 2340
Db 2292 GACACAATGGACGTAGCGCGCGTGTGGAGAGCTCTCTGGGCGGCGCAATGGACAGTCAAG 2351
Qy 2341 TGGATCCCGCACGACAAATCGTGA 2364
Db 2352 TGGATCCCGCACGACAAATCGTGA 2375

RESULT 11

ABK84353
ID ABK84353 standard; cDNA; 2716 BP.

XX
AC ABK84353;

XX
DT 14-AUG-2002 (first entry)

XX
DE Human cDNA differentially expressed in granulocytic cells #924.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;

KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US030821.

XX 03-OCT-2000; 2000US-0237189P.

XX (GENE-) GENE LOGIC INC.

XX Beazer-Barclay V, Weissman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression of
 PT genes associated with granulocyte activation, which serves as diagnostic
 PT markers that is useful for monitoring disease states and drug toxicity.

PS Claim 1; SEQ ID NO 924; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing the
 CC expression level to an expression level in an unactivated GC, where
 CC differential expression of Gs is indicative of GCA. Also included are
 CC modulating (M2) GA by contacting GC with an agent that alters the
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent
 CC capable of modulating GCA or an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease using the gene expression
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease, by detecting the level of
 CC expression in a sample of the tissue of gene(s) from Gs, where the level
 CC of expression of the gene is indicative of inflammation; (4) treating
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease, by contacting a tissue having inflammation with an
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
 CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
 CC for screening an agent capable of modulating GCA preferably in an
 CC inflammation in a tissue; M4 is useful for detecting an inflammation
 CC (especially chronic) in a tissue, an allergic response in a subject,
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
 CC disease, ulcerative colitis, periodontal disease; also bacterial
 CC infection, viral infection, parasitic infection, protozoal infection,
 CC fungal infection and M5 is useful for treating one of the above
 CC conditions. The present sequence represents a gene differentially
 CC expressed in granulocytes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2716 BP; 662 A; 707 C; 760 G; 587 T; 0 U; 0 Other;

Query Match 99.7%; Score 2356; DB 6; Length 2716;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2359; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGGCTGTGGATACAGCTCAGCAGCTCCAGAGAGAGCCCTTCATCAGATCAAGCG 60

Db	12	ATGGCTGTGGATACAGCTCAGCAGCTCCAGAGAGAGCCCTTCATCAGATCAAGCG	71
Qy	61	TTATATGGCCAGCATTTTCCATTGAGTGGGGCATTATTATCCAGTGGATTGAAGC	120
Db	72	TTATATGGCCAGCATTTTCCATTGAGTGGGGCATTATTATCCAGTGGATTGAAGC	131
Qy	121	CAAGCATGGGACTCAGTAGATCTTGATAATCCACAGGAGAACTTAAGAGCCACCCAGCTC	180
Db	132	CAAGCATGGGACTCAGTAGATCTTGATAATCCACAGGAGAACTTAAGAGCCACCCAGCTC	191
Qy	181	CTGAGGGGCTGTGTGAGGAGCTGCAGAGAGGACAGACACAGGTCACAGACACATATGAGG	240
Db	192	CTGAGGGGCTGTGTGAGGAGCTGCAGAGAGGACAGACACAGGTCACAGACACATATGAGG	251
Qy	241	TTTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGACACATATGAGG	300
Db	252	TTTTTACTGAAGATCAAGCTGGGGCACTATGCCACACAGCTCCAGACACATATGAGG	311
Qy	301	TGCCCATGGAGCTGGTCCGCTGCATCCGCAATATATTGTACATGAACAGAGGTTGGTC	360
Db	312	TGCCCATGGAGCTGGTCCGCTGCATCCGCAATATATTGTACATGAACAGAGGTTGGTC	371
Qy	361	CGAGAACCAACAAATGTAGCTCTCCAGCTGGAGGCTTGTGTATGTCATGTCACAGAAA	420
Db	372	CGAGAACCAACAAATGTAGCTCTCCAGCTGGAGGCTTGTGTATGTCATGTCACAGAAA	431
Qy	421	CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGAGCTGGAGTGGAGGACACAGAGAA	480
Db	432	CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGAGCTGGAGTGGAGGACACAGAGAA	491
Qy	481	GAGTAAAAAAGCTGCAGCAGACTCAGAGTACTTTCATCATCCAGTACAGAGAGCCCTG	540
Db	492	GAGTAAAAAAGCTGCAGCAGACTCAGAGTACTTTCATCATCCAGTACAGAGAGCCCTG	551
Qy	541	AGGATCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCAGAGGAGCTCTCAGCGG	600
Db	552	AGGATCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCCCAGAGGAGCTCTCAGCGG	611
Qy	601	GAGAGGCGCTCCAGCAGAGAGCAGTGTCTCTGGAGGCTGGTTGACAGCTGAGGACAG	660
Db	612	GAGAGGCGCTCCAGCAGAGAGCAGTGTCTCTGGAGGCTGGTTGACAGCTGAGGACAG	671
Qy	661	ACACTGCAGCAGTACCCGCTGGAGCTGCCGAGAGACACAGAGAGACCTTCAGCTGCTG	720
Db	672	ACACTGCAGCAGTACCCGCTGGAGCTGCCGAGAGACACAGAGAGACCTTCAGCTGCTG	731
Qy	721	CGGAAGCAGCAGACCATCATCTCTGGATGACGAGCTGATCCAGTGGAGCGGCGGACAG	780
Db	732	CGGAAGCAGCAGACCATCATCTCTGGATGACGAGCTGATCCAGTGGAGCGGCGGACAG	791
Qy	781	CTGGCCGGGAACGGCGGGCCCCCGAGGCGAGCTGAGAGCTGCTACAGTCTCTGCTGAG	840
Db	792	CTGGCCGGGAACGGCGGGCCCCCGAGGCGAGCTGAGAGCTGCTACAGTCTCTGCTGAG	851
Qy	841	AAGTTGGCGAGATCATCTCTGGAGAACCGCAGCAGATCCGAGGCTGAGACACTCTGC	900
Db	852	AAGTTGGCGAGATCATCTCTGGAGAACCGCAGCAGATCCGAGGCTGAGACACTCTGC	911
Qy	901	CAGCAGCTGCCATCCCGGCCCGCAGTGGAGGAGATGCTGGCCGAGGTCACAGCCACCATC	960
Db	912	CAGCAGCTGCCATCCCGGCCCGCAGTGGAGGAGATGCTGGCCGAGGTCACAGCCACCATC	971
Qy	961	ACGAGCATTTATCTCAGCCCTGGTGACAGCAGCTTCATCATTTGAGAGAGCCCTCTCAG	1020
Db	972	ACGAGCATTTATCTCAGCCCTGGTGACAGCAGCTTCATCATTTGAGAGAGCCCTCTCAG	1031
Qy	1021	GTCTGAAGACCCAGACCAAGTTTGGAGCCACTGTGCGCTGCTGGTGGGCGGGAAGCTG	1080
Db	1032	GTCTGAAGACCCAGACCAAGTTTGGAGCCACTGTGCGCTGCTGGTGGGCGGGAAGCTG	1091
Qy	1081	AAGCTGCACATGAACCCCGCCCGAGGTGAAGGCCACCATCATCAGTGAAGCAGCGCCCAAG	1140

Db 1092 AACGTGCACATGAACCCCTCCAGGTGAAGCCACCATCATCAGTGAGCAGCGGCCAAG 1151
Qy 1141 TCCTCTGCTCAAGAACGAGAACCCCGCAATGATTACAGTGGCGAGATCTTTGAACAATGTC 1200
Db 1152 TCCTCTGCTCAAGAACGAGAACCCCGCAATGATTACAGTGGCGAGATCTTTGAACAATGTC 1211
Qy 1201 TGGGTCTATGAGTAGTACCACCAAGCCACAGGCACCTTAGTGCCCACTTCACGAAATATGTCC 1260
Db 1212 TGGGTCTATGAGTAGTACCACCAAGCCACAGGCACCTTAGTGCCCACTTCACGAAATATGTCC 1271
Qy 1261 CTGAAACGAATTAAGAGGTGAGACCGTCTGTTGGGCGCAGAGTCGGTGACAGAGAAGAAAATTT 1320
Db 1272 CTGAAACGAATTAAGAGGTGAGACCGTCTGTTGGGCGCAGAGTCGGTGACAGAGAAGAAAATTT 1331
Qy 1321 ACAATCTCTTTGAAATCCCAAGTTGAGTGGTGGGAAATGAGCTGGTTTTCAAGTCAAG 1380
Db 1332 ACAATCTCTTTGAAATCCCAAGTTGAGTGGTGGGAAATGAGCTGGTTTTCAAGTCAAG 1391
Qy 1381 ACCCTGTCCTGCCAGTGGTGGTGGTTCATGGCAGCCAGGACCAACAATGCGAGCGCC 1440
Db 1392 ACCCTGTCCTGCCAGTGGTGGTGGTTCATGGCAGCCAGGACCAACAATGCGAGCGCC 1451
Qy 1441 ACTGTTCTCTGGGCAATGCTTTTGCAGAGCCTGGCAGGGTGCCATTTCGCGTCCCTGAC 1500
Db 1452 ACTGTTCTCTGGGCAATGCTTTTGCAGAGCCTGGCAGGGTGCCATTTCGCGTCCCTGAC 1511
Qy 1501 AAAGTGTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAATTCAGGCGCGAAGTGCAG 1560
Db 1512 AAAGTGTGTGGCCACAGCTGTGTGAGGCGCTCAACATGAATTCAGGCGCGAAGTGCAG 1571
Qy 1561 AGCAACCGGCGCTGACCAAGGAGAACCTCGTGTCTGCTGGCGAGAAACTGTTCAACAAC 1620
Db 1572 AGCAACCGGCGCTGACCAAGGAGAACCTCGTGTCTGCTGGCGAGAAACTGTTCAACAAC 1631
Qy 1621 AGCAGCAGCCACCTGGAGGACTACAGTGGCCTGTGTGTCTGCTGCCAGTTCACACAGG 1680
Db 1632 AGCAGCAGCCACCTGGAGGACTACAGTGGCCTGTGTGTCTGCTGCCAGTTCACACAGG 1691
Qy 1681 GAGAAATTACAGACCGGAATTACACTTTCTGGAATGGTTTGAACGTGTGATGGAAGTG 1740
Db 1692 GAGAAATTACAGACCGGAATTACACTTTCTGGAATGGTTTGAACGTGTGATGGAAGTG 1751
Qy 1741 TTTAAAAAACAATCTCAAGCCCTCATTTGAATGATGGGGCCATTTTGGGGTTTGTAAACAAG 1800
Db 1752 TTTAAAAAACAATCTCAAGCCCTCATTTGAATGATGGGGCCATTTTGGGGTTTGTAAACAAG 1811
Qy 1801 CAACAGGCCCCATGACCTACTGATTAAACAAGCCAGATGGGACCTTCCTCTGAGATTTCAGT 1860
Db 1812 CAACAGGCCCCATGACCTACTGATTAAACAAGCCAGATGGGACCTTCCTCTGAGATTTCAGT 1871
Qy 1861 GACTCAGAAATTGGCGGCATCACCATTTGCTTGGAGTTTGAATCTCAGGAAAGAAATGTTT 1920
Db 1872 GACTCAGAAATTGGCGGCATCACCATTTGCTTGGAGTTTGAATCTCAGGAAAGAAATGTTT 1931
Qy 1921 TGGAACTGATGCTTTTACACACAGAGCTTCTCCATCAGTCCCTPAGCCGCGCTTG 1980
Db 1932 TGGAACTGATGCTTTTACACACAGAGCTTCTCCATCAGTCCCTPAGCCGCGCTTG 1991
Qy 1981 GGAGACTTGAATTAACCTTATCTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATACTCC 2040
Db 1992 GGAGACTTGAATTAACCTTATCTATCTAGTGTTCCTGATCGGCCAAAAGATGAAGTATACTCC 2051
Qy 2041 AAATACTACACACAGTTCCCTGGAGTCTGCTACTCTAAAGCTGTGTGATGATGATCGTG 2100
Db 2052 AAATACTACACACAGTTCCCTGGAGTCTGCTACTCTAAAGCTGTGTGATGATGATCGTG 2111
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTGTGAAACGATCTGCAGATGCCGGGGGC 2160
Db 2112 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTGTGAAACGATCTGCAGATGCCGGGGGC 2171
Qy 2161 GGCAGCGCCAGTACATGGAACAGGCCCCCTCCACGCTGTGTGTCCCGCAGGCTCACTAT 2220
Db 2172 GGCAGCGCCAGTACATGGAACAGGCCCCCTCCACGCTGTGTGTCCCGCAGGCTCACTAT 2231

Qy 2221 AACATGTATCCACAGAACCTGACTCAGTCTTTGACACCGATGGGACTTCGATCTGGAG 2280
Db 2232 AACATGTATCCACAGAACCTGACTCAGTCTTTGACACCGATGGGACTTCGATCTGGAG 2291
Qy 2281 GACCAATGGACGTAGCGGGCGTGTGGAGGAGCTCTGTGGCCGGGCAATGGACAGTTCAG 2340
Db 2292 GACCAATGGACGTAGCGGGCGTGTGGAGGAGCTCTGTGGCCGGGCAATGGACAGTTCAG 2351
Qy 2341 TGGATCCCGCAGCAGCAATCGTGA 2364
Db 2352 TGGATCCCGCAGCAGCAATCGTGA 2375

RESULT 12

ADM79353
ID ADM79353 standard; cDNA; 2779 BP.

XX AC ADM79353;

XX AC ADM79353;

DT 03-JUN-2004 (first entry)

XX Human lymphoma associated, LA, contig #3.

XX ss: lymphoma associated protein; LAP; GNAS1; Pik3r1; G protein XI_{as};
XX NESP55; HIPK1; JAK1; neurogranin; Nrf2; lymphoma; leukaemia;
XX gene therapy; DNA vaccine; wound; inflammation; human;
XX expressed sequence tag; EST.

XX Homo sapiens.

XX US2003224460-A1.

XX 04-DEC-2003.

XX 24-SEP-2001; 2001US-00963131.

XX 22-SEP-2000; 2000US-00668644.

XX 13-JUL-2001; 2001US-00905390.

XX 13-JUL-2001; 2001US-00905491.

XX (PEDE/) PEDERSEN F S.

XX (SORE/) SORENSEN A B.

XX (HERN/) HERNANDEZ J M.

XX (NIEL/) NIELSEN A A.

XX (MOVI/) MOVING H.

XX Pedersen FS, Sorensen AB, Hernandez JM, Nielsen AA, Moving H;

XX WPI; 2004-033970/03.

XX Novel recombinant protein comprising lymphoma associated protein, useful
in treating lymphoma and leukemia.

XX Disclosure; SEQ ID NO 158; 99pp; English.

XX The invention relates to a recombinant protein comprising lymphoma
associated protein (LAP), chosen from a fully defined amino acid sequence
of human GNAS1, mouse Pik3r1, human Pik3r1, mouse G protein XI_{as}, human
G protein XI_{as}, mouse NESP55, human NESP55, mouse GNAS1, mouse HIPK1,
human HIPK1, mouse JAK1, human JAK1, mouse neurogranin, human
neurogranin, mouse Nrf2 and human Nrf2. LAP is useful for screening a
bioactive agent capable of binding to LAP. LAP is useful for screening a
bioactive agent capable for modulating the activity of LAP. Lymphoma
associated (LA) gene is useful for diagnosing lymphoma. LAP is useful for
treating lymphoma. LA gene is useful for diagnosing lymphomas or a
propensity to lymphomas. LA gene is useful for determining LA gene copy
number. LA genes are used as probe to determine chromosome location of
the LA gene. LA gene and LAP are useful in treating lymphoma and
leukaemia. LA gene is useful in gene therapy, useful as DNA vaccine. LA
gene is useful in generating animal models of lymphoma, which is useful
in screening bioactive molecules to treat lymphoma. LAP is useful in
treating wounds and inflammation. LAP is useful to generate polyclonal

CC and monoclonal antibodies. The present sequence represents a human
CC lymphoma associated, LA, contig.

SQ Sequence 2779 BP; 675 A; 735 C; 801 G; 559 T; 0 U; 9 Other;

Query Match 97.9%; Score 213.2; DB 12; Length 2779;

Best Local Similarity 99.3%; Pred. No. 0;

Matches 2349; Conservative 1; Mismatches 12; Indels 4; Gaps 3;

QY 1 ATGGCTGTGGATACAGCTCAGCAGCTCCAAAGGAGAGCCCTTCATCAGATCGACGG 60

DB 147 ATGGCTGTGTGGATACAGCTCAGCAGCTCCAAAGGAGAGCCCTTCATCAGATCGACGG 206

QY 61 TTATATGCCAGCATTTTCCCATTCAGGTGGGCAATTTATCCCATGGATTG--AAA 118

DB 207 TTATATGCCAGCATTTTCCCATTCAGGTGGGCAATTTATCCCATGGATTGAAA 265

QY 119 GCCAAGCATGGAGCTCAGTAGATCTTGATATATCCACAGGAGAAATTAAGGCCACCCAGC 178

DB 266 GCCAAGCATGGAGCTCAGTAGATCTTGATATATCCACAGGAGAAATTAAGGCCACCCAGC 324

QY 179 TCCTGGAGGCTGTGTCAGAGCTGCGAGAGCTGCAGAGAGGACAGCAGGTGGGAGATG 238

DB 325 TCCTGGAGGCTGTGTCAGAGCTGCGAGAGCTGCAGAGAGGACAGCAGGTGGGAGATG 384

QY 239 GGTTTTACTGAAGATCAAGCTGGGGCACTATGCCACAGCTCCAGAAACAGCTATGACC 298

DB 385 GGTTTTACTGAAGATCAAGCTGGGGCACTATGCCACAGCTCCAGAAACAGCTATGACC 444

QY 299 GCTGCCCATGGAGCTGCTCGCTGCATCCGCCATATATTGTACAAATGAACAGAGTTGG 358

DB 445 GCTGCCCATGGAGCTGCTCGCTGCATCCGCCATATATTGTACAAATGAACAGAGTTGG 504

QY 359 TCCGAGAGCCACAAATGGTAGCTCCAGCTGAGAGCTTGTGATGCCATGTCCACAGA 418

DB 505 TCCGAGAGCCACAAATGGTAGCTCCAGCTGAGAGCTTGTGATGCCATGTCCACAGA 564

QY 419 AACACCTCCAGATCAACAGAGCTTTGAGGAGCTGCGACTGTCAAGCAGGACACAGAGA 478

DB 565 AACACCTCCAGATCAACAGAGCTTTGAGGAGCTGCGACTGTCAAGCAGGACACAGAGA 624

QY 479 ATGAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTCATCATCCAGTACCAGGAGAGCC 538

DB 625 ATGAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTCATCATCCAGTACCAGGAGAGCC 684

QY 539 TGAGGATCCAAAGCTCAGTTTGGCCGCTGGCCAGCTGAGCCCCCAGGAGGCTGTAGCC 598

DB 685 TGAGGATCCAAAGCTCAGTTTGGCCGCTGGCCAGCTGAGCCCCCAGGAGGCTGTAGCC 744

QY 599 GGGAGAGCGCCCTCCAGCAGAGCAGGTGTCTCTGGAGGCTGTTGACGCTGAGGCAC 658

DB 745 GGGAGAGCGCCCTCCAGCAGAGCAGGTGTCTCTGGAGGCTGTTGACGCTGAGGCAC 804

QY 659 AGACACTGCAGCAGTACCGCTGGAGCTGCCGAGAGACACAGAAACCTTCGAGCTGC 718

DB 805 AGACACTGCAGCAGTACCGCTGGAGCTGCCGAGAGACACAGAAACCTTCGAGCTGC 864

QY 719 TCGGAGCAGCAGACCATCATCTCTGGATGACGAGCTGATCCAGTGGAGCGCGGAGC 778

DB 865 TCGGAGCAGCAGACCATCATCTCTGGATGACGAGCTGATCCAGTGGAGCGCGGAGC 924

QY 779 AGCTGGCCGGGAAACGGCGGCCCCCGAGGAGCGCTGAGTGCATCAGTCTGTGTG 838

DB 925 AGCTGGCCGGGAAACGGCGGCCCCCGAGGAGCGCTGAGTGCATCAGTCTGTGTG 984

QY 839 AGAAGTTGGCGGAGATCATCTGGCAGAAACCGGACAGCATCCGAGGCTGAGCACTCT 898

DB 985 AGAAGTTGGCGGAGATCATCTGGCAGAAACCGGACAGCATCCGAGGCTGAGCACTCT 1044

QY 899 GCCAGCAGCTGCCATCCCGGCCCGCAGTGGAGGAGATCTGGCGAGGTCAACGCCACCA 958

DB 1045 GCCAGCAGCTGCCATCCCGGCCCGCAGTGGAGGAGATCTGGCGAGGTCAACGCCACCA 1104

QY 959 TCACGACATTTATCTCAGCCCTCGTGTGACCAAGCAGCTTCATCATTTGAGAAGCAGCTCTCTC 1018

DB 1105 TCACGACATTTATCTCAGCCCTCGTGTGACCAAGCAGCTTCATCATTTGAGAAGCAGCTCTCTC 1164

QY 1019 AGGTCTGTGAGACCCACGACCAAGTTTTCAGCCACTGTGCGCTGTGCTGTGGCGGGAAGC 1078

DB 1165 AGGTCTGTGAGACCCACGACCAAGTTTTCAGCCACTGTGCGCTGTGCTGTGGCGGGAAGC 1224

QY 1079 TGAACGTGCACATGAAACCCGCCAGGTGAAGGCCACCATCATCAGTGAGCAGCAGGCA 1138

DB 1225 TGAACGTGCACATGAAACCCGCCAGGTGAAGGCCACCATCATCAGTGAGCAGCAGGCA 1284

QY 1139 AGTCTCTGCTCAAGAACGAGAACACCCGCCAATGATTACAGTGGGAGATCTTGAACAACT 1198

DB 1285 AGTCTCTGCTCAAGAACGAGAACACCCGCCAATGATTACAGTGGGAGATCTTGAACAACT 1344

QY 1199 GCTGCGTCATGGAGTACCAAGCCACAGCAGCCCTTAGTGCCACTTTCAGGAATATGT 1258

DB 1345 GCTGCGTCATGGAGTACCAAGCCACAGCAGCCCTTAGTGCCACTTTCAGGAATATGT 1404

QY 1259 CCCTGAAACGAATTAAGAGGTTCAGACCGTCTGTGGGCGAGTCCGTGACAGAAAGAAAT 1318

DB 1405 CCCTGAAACGAATTAAGAGGTTCAGACCGTCTGTGGGCGAGTCCGTGACAGAAAGAAAT 1464

QY 1319 TTACAATCCTGTGTTGAATCCAGTTCAGTGTGTTGTTGAAATTCAGTGTGTTTCAAGTCA 1378

DB 1465 TTACAATCCTGTGTTGAATCCAGTTCAGTGTGTTGTTGAAATTCAGTGTGTTTCAAGTCA 1524

QY 1379 AGACCTGTCTCTGCCAGTGTGTGATCGTTCATGTCAGCCAGGACCAACATTCGACGG 1438

DB 1525 AGACCTGTCTCTGCCAGTGTGTGATCGTTCATGTCAGCCAGGACCAACATTCGACGG 1584

QY 1439 CCAGTGTCTCTGGGCAATGCTTTTCAGAGCTTCGAGAGCTTCGAGGCTGTCCTGCTGCTG 1498

DB 1585 CCAGTGTCTCTGGGCAATGCTTTTCAGAGCTTCGAGAGCTTCGAGGCTGTCCTGCTGCTG 1644

QY 1499 ACAAGTGTCTGTGCCACAGCTGTGTGAGGCGCTCAACATGAAATTCAGGCGCCAGTGC 1558

DB 1645 ACAAGTGTCTGTGCCACAGCTGTGTGAGGCGCTCAACATGAAATTCAGGCGCCAGTGC 1704

QY 1559 AGACAAACCGGGGCTGACCAAGGAGAACTCTGTTCTCTGGCGCAGAAAACCTGTTCAACA 1618

DB 1705 AGACAAACCGGGGCTGACCAAGGAGAACTCTGTTCTCTGGCGCAGAAAACCTGTTCAACA 1764

QY 1619 ACAGCAGAGCCACTTCGAGGACTACAGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1678

DB 1765 ACAGCAGAGCCACTTCGAGGACTACAGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1824

QY 1679 GGGAGAAATTTACAGGACGGAATTCACATTTCTGGCAATGTTGACGCTGTGATGGAAG 1738

DB 1825 GGGAGAAATTTACAGGACGGAATTCACATTTCTGGCAATGTTGACGCTGTGATGGAAG 1884

QY 1739 TGTAAAAAACAATCTCAAGCCTCATTTGGAATGATGGGCGCATTTTGGGGTTTGTAAAAA 1798

DB 1885 TGTAAAAAACAATCTCAAGCCTCATTTGGAATGATGGGCGCATTTTGGGGTTTGTAAAAA 1944

QY 1799 AGCAACAGGCGCCATGACCTACTGATTAACAGCAGATGGACCTTCTCTCTGAGATTC 1858

DB 1945 AGCAACAGGCGCCATGACCTACTGATTAACAGCAGATGGACCTTCTCTCTGAGATTC 2004

QY 1859 GTGACTCAGAAATTTGGCGGATCACCATTTGCTGGAAAGTTGATTTCTCAGGAAAGAAATGT 1918

DB 2005 GTGACTCAGAAATTTGGCGGATCACCATTTGCTGGAAAGTTGATTTCTCAGGAAAGAAATGT 2064

QY 1919 TTTGGAATCTGATGCTTTTACCAACAGAGATTTCTCCATCAGGTCCCTAGCCGACGCT 1978

DB 2065 TTTGGAATCTGATGCTTTTACCAACAGAGACTTCTCCATCAGGTCCCTAGCCGACGCT 2124

QY 1979 TGGAGACTTGAATTAATCTTATCTAAGTGTCTGATCGGCCAAAGAGATGAATGAT 2038

DB 2125 TGGAGACTTGAATTAATCTTATCTAAGTGTCTGATCGGCCAAAGAGATGAATGAT 2184

QY 2039 CCAATACTACACACAGTTCCTTCGCGAGTCTGCTACTGCTAAAGCTGTTGATGGATAGC 2098

Db	1921	GCAGTGGTCCCTGAGTTTGTAACGATCTGCAGATGTCGGGGGGGGGAGCCGACGTA	1980
Qy	2175	CATGGACAGGCCCCCTCCACGCTGTGTCTCCCGAGGCTCACTATAAATGTACCCACA	2234
Db	1981	CATGGACAGGCCCCCTCCACGCTGTGTCTCCCGAGGCTCACTATAAATGTACCCACA	2040
Qy	2235	GAACCTGACTAGTCTTGTGACACCGATGGGACTTTCGATCTGGAGGACACAATGGACGT	2294
Db	2041	GAACCTGACTAGTCTTGTGACACCGATGGGACTTTCGATCTGGAGGACACAATGGACGT	2100
Qy	2295	AGCGGCGGTGCGAGGAGCTCTCGGCGCGGCAATGGACAGTCAGTGGATCCGCAACG	2354
Db	2101	AGCGGCGGTGCGAGGAGCTCTCGGCGCGGCAATGGACAGTCAGTGGATCCGCAACG	2160
Qy	2355	ACAATCGTGA	2364
Db	2161	ACAATCGTGA	2170
RESULT 14			
ACN40207	ID	ACN40207 standard; cDNA; 2776 BP.	
XX	AC	ACN40207;	
XX	DT	18-NOV-2004 (first entry)	
XX	DE	Tumour-associated antigenic target (TAT) cDNA DNA326348, SEQ ID NO:4849.	
XX	KW	Tumour-associated antigenic target; TAT; human; overexpression; cancer;	
XX	KW	tumour; diagnosis; cell proliferative disorder; breast cancer;	
XX	KW	colorectal cancer; lung cancer; ovarian cancer; liver cancer;	
XX	KW	central nervous system cancer; bladder cancer; pancreatic cancer;	
XX	KW	cervical cancer; melanoma; leukaemia; hybridisation probe;	
XX	KW	chromosome identification; chromosome mapping; gene mapping;	
XX	KW	gene therapy; cytostatic; gene; ss.	
OS	XX	Homo sapiens.	
PN	XX	WO2004030615-A2.	
XX	XX	15-APR-2004.	
XX	XX	29-SEP-2003; 2003WO-US028547.	
XX	XX	02-OCT-2002; 2002US-0414971P.	
PR	XX	(GETH) GENENTECH INC.	
PA	XX	Wu TD, Zhang Z, Zhou Y;	
PI	XX	WPI; 2004-347921/32.	
DR	XX	New tumor-associated antigenic target polypeptides and nucleic acids,	
PT	XX	useful in preparing a medicament for treating or detecting a	
PT	XX	proliferative disorder, e.g. breast, lung, colorectal, ovarian or	
PT	XX	prostate cancer or tumor.	
XX	XX	Claim 1; SEQ ID NO 4849; 7273pp; English.	
XX	XX	The invention relates to human tumour-associated antigenic target (TAT)	
XX	XX	polypeptides, and their related nucleic acids. The TAT polypeptides are	
CC	XX	overexpressed in cancer tissues compared to normal tissues, and may thus	
CC	XX	serve as effective targets for the diagnosis and treatment of cancer in	
CC	XX	mammals. The invention also relates to nucleic acid and polypeptide	
CC	XX	sequences at least 80% identical to the TAT nucleic acids and	
CC	XX	polypeptides; expression vectors and host cells comprising a TAT nucleic	
CC	XX	acid; an antibody specific for a TAT polypeptide; a peptide or organic	
CC	XX	molecule which binds to a TAT polypeptide; fusion proteins comprising a	
CC	XX	TAT polypeptide; and methods and compositions for the treatment or	
CC	XX	diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,	
CC	XX	antibodies, antagonists, binding molecules and compositions are useful	
CC	XX	for diagnosing or treating a cell proliferative disorder associated with	

CC increased TAR expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAR nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
XX represents a TAR nucleic acid of the invention
SQ

Sequence 2776 BP; 607 A; 684 C; 734 G; 513 T; 0 U; 238 Other;

Query Match 87.7%; Score 2073.2; DB 13; Length 2776;

Best Local Similarity 88.9%; Pred. No. 0;

Matches 2101; Conservative 0; Mismatches 256; Indels 7; Gaps 1;

Qy	1	ATGGCTGTGGATACAAGCTCAGACCTCGAGGAGAGCCCTTCATCAGATGCAAGCG	60
Db	148	ATGGCTGTGGATACAAGCTCAGACCTCGAGGAGAGCCCTTCATCAGATGCAAGCG	207
Qy	61	TTATATGCCAGCATTTTCCATTGAGGTGGGCATTATTATCCCAAGTGGATTGAAAGC	120
Db	208	TTATATGCCAGCATTTTCCATTGAGGTGGGCATTATTATCCCAAGTGGATTGAAAGC	267
Qy	121	CAAGCATGGGACTCAGTAGATCTTTGATAATCCACAGAGAACATTAAGGCCACCCAGCTC	180
Db	268	CAAGCATGGGACTCAGTAGATCTTTGATAATCCACAGAGAACATTAAGGCCACCCAGCTC	327
Qy	181	CTGAGGGCCCTGTGTCAGGAGCTGCAGAAAGGCAGACACAGGTGGGGGAAGATGGG	240
Db	328	CTGAGGGCCCTGTGTCAGGAGCTGCAGAAAGGCAGACACAGGTGGGGGAAGATGGG	387
Qy	241	TTTTTACTGAAGATCAAGCTGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC	300
Db	388	TTTTTACTGAAGATCAAGCTGGGCACTATGCCACACAGCTCCAGAACACGTATGACCGC	447
Qy	301	TGCCCCATGGAGCTGGTCCGCTGCATCCGCCATATATTGTACAAATGAACAGAGTTGGTC	360
Db	448	TGCCCCATGGAGCTGGTCCGCTGCATCCGCCATATATTGTACAAATGAACAGAGTTGGTC	507
Qy	361	CGAGAACCAACATGTTAGTCTCCAGCTGGAAGCCTTGGTATGCCATGTCCAGAAA	420
Db	508	CGAGAACCAACATGTTAGTCTCCAGCTGGAAGCCTTGGTATGCCATGTCCAGAAA	567
Qy	421	CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGCACTGGTTCAGCAGGACACAGAGAT	480
Db	568	CACCTCCAGATCAACACAGAGCTTTGAGGAGCTGGCACTGGTTCAGCAGGACACAGAGAT	627
Qy	481	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATATCCAGTACCGAGAGCCTG	540
Db	628	GAGTTAAAAAGCTGCAGCAGACTCAGGAGTACTTTCATATCCAGTACCGAGAGCCTG	687
Qy	541	AGGATCCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCGCCAGGAGCGTCTGAGCCGG	600
Db	688	AGGATCCAAGCTCAGTTTGGCCCGCTGGCCAGCTGAGCCGCCAGGAGCGTCTGAGCCGG	747
Qy	601	GAGACGCGCTCCACAGCAGAGCAGTCTCTCTGAGGCGCTGGTTGACGCTGAGGCACAG	660
Db	748	GAGACGCGCTCCACAGCAGAGCAGTCTCTCTGAGGCGCTGGTTGACGCTGAGGCACAG	807
Qy	661	ACACTGCAGCAGTACCGCGTGGAGCTGCCGAGAGCACCAGAAAGACCCCTGCAGCTGCTG	720
Db	808	ACACTGCAGCAGTACCGCGTGGAGCTGCCGAGAGCACCAGAAAGACCCCTGCAGCTGCTG	867
Qy	721	CGGAAGCAGCAGCAGCATCTGGATGACGAGCTGATCCAGTGGAAAGCGCGGCAGCAG	780
Db	868	CGGAAGCAGCAGCAGCATCTGGATGACGAGCTGATCCAGTGGAAAGCGCGGCAGCAG	927
Qy	781	CTGCGCGGGAAACGCGCGCCCCCGAGCGAGCCTGGACGTGCTACAGTCTGGTGTGAG	840
Db	928	CTGCGCGGGAAACGCGCGCCCCCGAGCGAGCCTGGACGTGCTACAGTCTGGTGTGAG	987
Qy	841	AAGTTGGCGGAGATCATCTGSCAGAACCGGCAGCAGATCCGAGGGGTGAGCACTCTGC	900
Db	988	AAGTTGGCGGAGATCATCTGSCAGAACCGGCAGCAGATCCGAGGGGTGAGCACTCTGC	1047

Qy	901	CAGCAGTGCCTATCCCGGCCCCAGTGGAGAGATGCTGSCGAGGTCAACGCCACCATC	960
Db	1048	CAGCAGTGCCTATCCCGGCCCCAGTGGAGAGATGCTGSCGAGGTCAACGCCACCATC	1107
Qy	961	ACGGACATATATCTCAGCCCTGGTGACGACACCTTTCATATTTGAGAAGCAGCCTCTCTCAG	1020
Db	1108	ACGGACATATATCTCAGCCCTGGTGACGACACCTTTCATATTTGAGAAGCAGCCTCTCTCAG	1167
Qy	1021	GTCTGTAAGACCCAGACCAAGTTTTCAGCCACTGTGGGCTGTCTGGTGGCGGGAAGCTG	1080
Db	1168	GTCTGTAAGACCCAGACCAAGTTTTCAGCCACTGTGGGCTGTCTGGTGGCGGGAAGCTG	1227
Qy	1081	AACGTGCACATGAACCCCGCCAGGTGAAGGCCACCATCATCAGTCAGCAGCAGGCCAAG	1140
Db	1228	AACGTGCACATGAATCCCCCGCAGGTGAAGGCCACCATCATCAGTCAGCAGGCCAAG	1287
Qy	1141	TCCTCTGTCAAGAACGAGAACACCCCGCAATGATATACAGTGGCGGAGATCTTTGAACAACATGC	1200
Db	1288	TCCTCTGTAAATAATGAGAACACCCCGCAANNNNNNNNNTGTTGAGATCTCTGAACAACATGC	1347
Qy	1201	TGGTCAATGGAGTACACCAAGCCACAGGCAACCTTTAGTCCCACTTCAGGAATATGTCC	1260
Db	1348	TGGTCAATGGAGTACACCAAGCCACCGGCAACCTTTAGTCCCACTTCAGGAATATGTCC	1400
Qy	1261	CTGAACGAAATTAAGAGTTCAGACCGTCTGGGGCAGAGTCGGTGAAGAGAAAAATTT	1320
Db	1401	NN	1460
Qy	1321	ACAATCCTGTTGAATCCCAAGTTTCAAGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG	1380
Db	1461	NN	1520
Qy	1381	ACCTGTCTCTGCGAGTGGTGGTTCATGCGCAGCCAGGACAAACAATGGCAAGGCC	1440
Db	1521	NN	1580
Qy	1441	ACTGTTCTCTGGGACAATGCTTTTGACAGAGCCTGGCAGGCTGCATTTGCGCTGCCTGAC	1500
Db	1581	NN	1640
Qy	1501	AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGGCGCGAAGTGCAG	1560
Db	1641	AAAGTGTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGGCGCGAAGTGCAG	1700
Qy	1561	AGCAACCGGGGCTGACCAAGGAGAACCTGCTGTTCTGCGCAGAGAACTGTTCAACAAC	1620
Db	1701	AGCAACCGGGGCTGACCAAGGAGAACCTGCTGTTCTGCGCAGAGAACTGTTCAACAAC	1760
Qy	1621	AGCAGCAGCCACTGGAGGACTACAGTGGCTCTGCTGTTCTGCTGGTCCCAAGTTCAACAGG	1680
Db	1761	AGCAGCAGCCACTGGAGGACTACAGTGGCTCTGCTGTTCTGCTGGTCCCAAGTTCAACAGG	1820
Qy	1681	GAGAAATTTACAGGAGCGAATTAACATTTCTGGCAATGGTTTGAACGCTGTGATGGAAGTG	1740
Db	1821	GAGAAATTTACAGGAGCGAATTAACATTTCTGGCAATGGTTTGAACGCTGTGATGGAAGTG	1880
Qy	1741	TTAAAAAACAATCTAAGCCTCATTTGGAATGATGGGCGCAATTTTGGGTTTGTAAAAAAG	1800
Db	1881	TTAAAAAACAATCTAAGCCTCATTTGGAATGATGGGCGCAATTTTGGGTTTGTAAAAAAG	1940
Qy	1801	CAACAGGCCCATCAGCTTACTGATTAAACAGCCAGATGGGACCTTCTCTCTGAGATTTCAGT	1860
Db	1941	CAACAGGCCCATCAGCTTACTGATTAAACAGCCAGATGGGACCTTCTCTCTGAGATTTCAGT	2000
Qy	1861	GACTCAGAAATTTGGCGGCATCACCAATTTGCTTTGGAAGTTTTCATTTCTCAGGAAAGATGTTT	1920
Db	2001	GACTCAGAAATTTGGCGGCATCACCAATTTGCTTTGGAAGTTTTCATTTCTCAGGAAAGATGTTT	2060
Qy	1921	TGGAATCTGATGCTTTTACCAACAGAGACTTCTCCATCAGGTCCTCTAGCCGACCGCTTG	1980
Db	2061	TGGAATCTGATGCTTTTACCAACAGAGACTTCTCCATCAGGTCCTCTAGCCGACCGCTTG	2120

QY 1981 GGAGACTTGAATTAACCTTATCTACGTGTTTCTGTATCGGCCCAAAAGATGAAGTATACCTCC 2040
DB 2121 GGAGACTTGAATTAACCTTATCTACGTGTTTCTGTATCGGCCCAAAAGATGAAGTATACCTCC 2180
QY 2041 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2100
DB 2181 AAATACTACACACAGTTCCCTGCGAGTCTGCTACTGCTAAAGCTGTTGATGGATACGTG 2240
QY 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGGC 2160
DB 2241 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTTTGTGAACGCATCTGCAGATGCCGGGGGC 2300
QY 2161 GGCAGCGCCAGTACATGCAACAGCCCTCCCGAGCTGTGTGTCCTCCAGCTCACTAT 2220
DB 2301 GGCAGCGCCAGTACATGCAACAGCCCTCCCGAGCTGTGTGTCCTCCAGCTCACTAT 2360
QY 2221 AACATGTACCCACAGAACCTGACTCAGTCTCTTGACACCGATGGGACTTCGATCTGGAG 2280
DB 2361 AACATGTACCCACAGAACCTGACTCAGTCTCTTGACACCGATGGGACTTCGATCTGGAG 2420
QY 2281 GACCAATGGAGCTAGCGCGCGGTGTGGAGGAGCTCTTGGCGCGCCCAATGGACAGTCCAG 2340
DB 2421 GACCAATGGAGCTAGCGCGCGGTGTGGAGGAGCTCTTGGCGCGCCCAATGGACAGTCCAG 2480
QY 2341 TGGATCCCGCAGCACAAATCGTGA 2364
DB 2481 TGGATCCCGCAGCACAAATCGTGA 2504

RESULT 15

ID ADB58436 standard; DNA; 2671 BP.

XX AC ADB58436;

XX DT 04-DEC-2003 (first entry)

XX DE Toxicity-related gene, SEQ ID 3462.

XX KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;

XX KW drug screening; toxicity assay; db.

XX OS Unidentified.

XX XX WO2003064624-A2.

XX XX 07-AUG-2003.

XX XX 31-JAN-2003; 2003WO-US003194.

XX XX 31-JAN-2002; 2002US-00060087.

XX XX 15-MAR-2002; 2002US-0364045P.

XX XX 15-MAR-2002; 2002US-0364055P.

XX XX 30-DEC-2002; 2002US-0436643P.

XX XX (GENE-) GENE LOGIC INC.

XX XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

XX XX WPI; 2003-689530/65.

XX XX Predicting a toxic effect of a compound, useful in identifying toxicity

XX XX markers in liver tissues or cells for drug screening and toxicity assays,

XX XX comprises preparing gene expression profile of tissue or cells exposed to

XX XX the compound.

XX XX Claim 1; SEQ ID NO 3462; 1156pp; English.

XX XX The present invention relates to a method for predicting a toxic effect

XX XX of a compound. The method comprises preparing a gene expression profile

XX XX of a tissue or cell sample exposed to the compound, and comparing the

XX XX gene expression profile to a database comprising SEQ ID 1-4925, where

XX XX differential expression of the gene indicates at least one toxic effect.

CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2671 BP; 638 A; 739 C; 776 G; 518 T; 0 U; 0 Other;

Query Match 84.2%; Score 1991.2; DB 10; Length 2671;

Best Local Similarity 90.4%; Pred. No. 0;

Matches 2138; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 1 ATGCTGTGTGGATACAAAGCTCAGCAGCTCAAGGAGAAGCCCTTCATCAGATGCAAGCG 60
DB 111 ATGCAATGTGGATACAGGCTCAGCAGCTCCAGGCGATGCCCTTCATCAGATGCAAGCG 170
QY 61 TTATATGGCAGCAATTTCCCATTTGAGGTGGCGCATTTATTTATCCAGTGGATTGAAGC 120
DB 171 TTGTACGCGCAGCATTTCCCATCGAGGTGCGACATTTATTTATCCAGTGGATTGAAGT 230
QY 121 CMAGCATGGGACTCAGTAGATCTTGATTAATCCAGGAGACATTTAAGGCCACCCAGCTC 180
DB 231 CAAGCCTGGGACTCAATAGATCTTTGATTAATCCAGGAGAACATTTAAGGCCACCCAGCTC 290
QY 181 CTGAGGCGCTGTGTGAGGAGCTCAGAGAAGGAGCAGCAGCTGGGGGAAGATGGG 240
DB 291 CTGGAGGCGCTGTGTGAGGAGCTCAGAGAAGGAGCAGCAGCTGGGGGAAGATGGG 350
QY 241 TTTTACTGAAGATCAAGCTGGGCACTATGCCACACAGCTCCAGAACACATATGACCGC 300
DB 351 TTTTGTGTAAGATCAAGCTGGGCACTATGCCACCGAGCTCCAGAACACATATGACCGC 410
QY 301 TGCCCATGGAGCTGGTCCGCTGCATCGGCATATATTGTACATGACAGAGTTGGTC 360
DB 411 TGCCCATGGAGCTGGTCCGCTGCATCGGCATATATTGTACATGACAGAGTTGGTC 470
QY 361 CGAGAAGCCAAACATGGTAGCTCTCCAGCTGGAAGCCCTTGCTGATGCTCCAGAAA 420
DB 471 CGAGAAGCCAAACATGGTAGCTCTCCGCTGGAAGCTTGTGTGAGCCATGTCCAGAA 530
QY 421 CACCTCCAGATCAACAGACGTTTGGAGAGCTGCGACTGGTCAACGAGGACACAGAGAA 480
DB 531 CACCTCCAGATCAACCAAAACGTTTGGAGAGCTGCGCTCTGATCAACAGGACACAGAGAG 590
QY 481 GAGTTAAAAGCTGCGAGCAGACTCAGAGTACTTTCATCCTCCAGTACCAGAGAGCCCTG 540
DB 591 GAGCTGAAGAGCTGCGAGCAGAGCCCAAGAGTACTTTCATCCTCCAGTACCAGAGAGCC 650
QY 541 AGGATCAAGCTCAGTTTGGCCCGCTGCGAGCTGAGCCCGAGGAGGCTCTCAGAGCGG 600
DB 651 CGGATCCAGCTCAGTTTGGCCCGCTGCGAGCTGAGCCCGAGGAGGCTCTCAGAGCGG 710
QY 601 GAGACGCGCTCCAGCAGAGAGAGGTTCTCTGTGGAGGCTGTGTTGTCAGCGCTGAGGCA 660
DB 711 GAGACGCGCTCCAGCAGAGAGAGGTTCTCTGTGGAGGCTGTGTTGTCAGCGCTGAGGCA 770
QY 661 ACATGAGCAGATACCCGCTGGAGCTCCCGAGAGAGCAGCAGAGAGCCCTCAGCTGCTG 720
DB 771 ACATGAGCAGATACCCGCTGGAGCTCCCGAGAGAGCAGCAGAGAGCCCTCAGCTGCTG 830
QY 721 CGGAAGCAGCAGACCATCATCTCGATGACAGCTGATCCAGTGGAGAGCGGCGCAGCAG 780
DB 831 AGGAAGCAGCAGACCATCATCTCGAGCAGAGCTGATCCAGTGGAGAGCGGCGCAGCAG 890
QY 781 CTGCGCGGGAACCGCGGCGCCCGCAGGCGAGCTGGAGCTGCTACAGTCTCTGCTGAG 840

Db 891 CTGCGCGGAATCGGGCTCTCTGAGGGCAGCCTGGATGCTGCGAGTCTCTGCTGTGAG 950
Qy 841 AAGTTGGCGGAGATCATCTGGCAGAAACCGGAGCAGATTCGCGAGGGCTGAGCACCTCTGC 900
Db 951 AAGTGGCCGAGATCATCTGGCAGAAACCGGAGCAGATTCGCGAGGGCTGAGCACCTCTGC 1010
Qy 901 CAGCAGCTGCCCATCCCGGCCAGTGGAGAGATCTGCCGAGGTCACAGCCACCATC 960
Db 1011 CAGCAGCTGCCCATCCCGGCCAGTGGAGAGATCTGCCGAGGTCACAGCCACCATC 1070
Qy 961 ACGACATATCTCAGGCCCTGGTACCAGCAGCATCTCATCTATGAGAGCAGCCCTCTCAG 1020
Db 1071 ACAGACATCATCTCAGCCCTGGTACCAGCAGCATCTCATCTCAGAGAGCAGCCCTCTCAG 1130
Qy 1021 GTCTGAAGACCCAGAACAAAGTTTGGAGCCACTGTGCGCTCTGTGGTGGCGGGAAGCTG 1080
Db 1131 GTCTGAAGACCCAGAACAAAGTTTGGAGCCACTGTGCGCTCTGTGGTGGCGGGAAGCTG 1190
Qy 1081 AACGTGCACATGAACCCCGCCAGGTGAAGCCACCATCATCATGAGAGCAGCGGCCAAG 1140
Db 1191 AACGTGCACATGAACCCCTCGCAGGTGAAGCCACCATCATCATGAGAGCAGCGGCCAAG 1250
Qy 1141 TCTCTGTCAAGAACGAGAACACCCGCAATGATTAAGTGGCGAGATCTTGAACAACTGC 1200
Db 1251 TCCCTGTCAAGAACGAGAACACCCGCAATGATTAAGTGGCGAGATCTTGAACAACTGC 1310
Qy 1201 TGCCTCATGAGATACCAACGACCAAGGCCACAGGCCCTTATAGTGCCTCATCTCAGGAATATGTC 1260
Db 1311 TGCCTCATGAGATACCAACGACCAAGGCCACAGGCCCTTATAGTGCCTCATCTCAGGAATATGTC 1370
Qy 1261 CTGAAACGAATTAGAGGTCAGACCGTCTGTGGGCGAGTCGGTGACAGAGAAATTT 1320
Db 1371 CTGAAACGAATTAGAGGTCAGACCGTCTGTGGGCGAGTCGGTGACAGAGAAATTT 1430
Qy 1321 ACAATCTGTTGAATCCCAAGTTCAGTGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1380
Db 1431 ACAATCTGTTGAATCCCAAGTTCAGTGTGGTGGAAATGAGCTGGTTTTTCAAGTCAAG 1490
Qy 1381 ACCGTGCTCCCTGCGAGTGGTGGTATCGTTTCATGGCGCCAGGACCAATGCGACGGCC 1440
Db 1491 ACCGTGCTCCCTGCGAGTGGTGGTATCGTTTCATGGCGCCAGGACCAATGCGACGGCT 1550
Qy 1441 ACTGTTCTCTGGGACAATGCTTTTGCAGAGCTGCGAGGGTGCCATTTGCGTGCCTGAC 1500
Db 1551 ACCGTCTCTGGGACAACGCTTTTGCAGAGCTGCGAGGGTGCCATTTGCGTGCCTGAC 1610
Qy 1501 AAAGTGTCTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGGGCCGAGTGCAG 1560
Db 1611 AAAGTGTCTGTGGCCACAGCTGTGTGAGCGCTCAACATGAAATTCAGGGCTGAGTGCAG 1670
Qy 1561 AGCAACCGGGGCTTGACCAAGGAGAACTCGTGTTCCTGGCGAGAACTGTTCAACAGC 1620
Db 1671 AGCAACCGGGGCTTGACCAAGGAGAACTCGTGTTCCTGGCGAGAACTGTTCAACAGC 1730
Qy 1621 AGCAGAGCCACCTGGAGGACTACAGTGGCTGTCTGTCTCGTCCAGTTCACAGG 1680
Db 1731 AGCAGAGCCACCTGGAGGACTACAGTGGCTGTCTGTCTCGTCCAGTTCACAGG 1790
Qy 1681 GAGAAATTTACAGGACGGAATTTACACTTTCTGGCAATGGTTTGAACGGTGTGATGGAAGTG 1740
Db 1791 GAGAAATTTACAGGACGGAATTTACACTTTCTGGCAATGGTTTGAACGGTGTGATGGAAGTG 1850
Qy 1741 TTAAGAAACATCTCAAGCCTCATTTGGAATGATGGGCGCATTTTGGGGTTTGTAAACAAG 1800
Db 1851 TTAAGAAACATCTCAAGCCTCATTTGGAATGATGGGCGCATTTTGGGGTTTGTAAACAAG 1910
Qy 1801 CAACAGCCCATGACTACTGATTAACAAGCAGATGGACCTTCTCTCCTGAGATTCAGT 1860
Db 1911 CAGCAGCCCATGACTACTGATTAACAAGCAGATGGACCTTCTCTCCTGAGATTCAGT 1970
Qy 1861 GACTCAGAAATTTGGCGGCATCACCATTGCTTTGGAAGTTTGTATCTCAGGAAGAATGTTT 1920
Db 1971 GACTCAGAAATTTGGCGGCATCACCATTGCTTTGGAAGTTTGTATCTCAGGAAGAATGTTT 2030

Qy 1921 TGAATCTGATGCTTTTACCACAGAGACTTCTCCATCAGGTCCCTAGCCGACCGCTTG 1980
Db 2031 TGAATCTGATGCTTTTACCACAGAGACTTCTCTATCCGGTCCCTCGCTGACCGCTG 2090
Qy 1981 GGAGACTTGAATTAACCTTATCTACGTGTTTCTGTATCGGCCAAAGATGAAGTATATCTC 2040
Db 2091 GGAGACTTGAATTAACCTTATCTATATACGTGTTTCTGTATCGGCCAAAGATGAAGTATATCT 2150
Qy 2041 AAATACTACACACAGTTCCTCGAGTCTGTCTAAAGCTTGTATGATGATAGTG 2100
Db 2151 AAGTACTACACACAGTTCCTCGTGTAGCCGCCACCGCGAAAGCAGTGCACGATACGTG 2210
Qy 2101 AAGCCACAGATCAAGCAAGTGGTCCCTGAGTGTGTGAACGATCTGCAGATGCCGGGC 2160
Db 2211 AAGCCCGCAGATCAAGCAAGTGGTCCCGAGTTTGTCAATGCTTCTGCAGATGACGGAGC 2270
Qy 2161 GGCAGCGCCACGTCATCATGGACAGGCCCTCCCGAGCTGTGTCTCCAGGCTCACTAT 2220
Db 2271 G---GGCCACCTACATGGACAGGCTCTTCCCGAGTGTGTGCCCTCAGGCTCACTAT 2327
Qy 2221 AACATGTACCCACAGAACCTGACTCAGTCTTTGACACCGATGGGACCTTCGATCTGGAG 2280
Db 2328 AACATGTACCCACAGAAATCCGACTCTGTCTCTGATACCGATGGGATTTTCGATCTGGAA 2387
Qy 2281 GACACAAATGGACGTAGCGCGCGTGTGGAGGAGCTCTTGGCGCGCCAAATGGACAGTCAG 2340
Db 2388 GACACAAATGGACGTGGACGCGCTGTGGAAGAGCTCTTAGGCCGGCCCATGGACAGTCAG 2447
Qy 2341 TGGATCCCGCACGCAACATCGTGA 2364
Db 2448 TGGATCCCTCACGGCAGTCATGA 2471

Search completed: May 17, 2005, 01:42:31
Job time : 1838 secs